VOL. XXVI JUNE, 1942 No. 3

THE JOURNAL OF APPLIED PSYCHOLOGY

JAMES P. PORTER, *Editor*
OHIO UNIVERSITY ATHENS, OHIO

With a Board of Cooperating Editors
and

PAUL S. ACHILLES HENRY C. LINK
Associate Editors Representing
THE PSYCHOLOGICAL CORPORATION

SOME INTERESTING CONTRIBUTIONS IN THIS ISSUE:

DANIEL STARCH. The 100 Greatest Books Selected by 100 Qualified Persons

EDWARD C. WEBSTER. A Follow-Up on Vocational Guidance

H. MEILTZER. Mental Hygiene and Learning

ROBERT J. LEWINSKI. A Qualitative Analysis of the Kent Oral Emergency Test as a Clinical Instrument in the Examination of Naval Recruits

PUBLISHED BI-MONTHLY BY THE EDITOR
OHIO UNIVERSITY, ATHENS, OHIO

PRINTED BY THE SCIENCE PRESS PRINTING COMPANY, LANCASTER, PA.

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry May 7, 1930, at the post office at Lancaster, Pa.
Copyright 1942 by James P. Porter.

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. Bingham, Personnel Research Federation, New York; M. E. Broom, State Teachers College, San Diego, Calif.; H. E. Burtt, Ohio State University; Rossiter H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; J. E. W. Wallin, Division of Special Education, Wilmington, Del.; Herbert S. Langfeld, Princeton University; Henry M. McComas, Johns Hopkins University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, New York; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; Joseph Jastrow, New School of Social Research, New York; G. M. Whipple, Danvers, Mass.; Helen Thompson Woolley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and offprints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by the Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $4.00 a year. Single numbers $1.25 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

Vol. XXVI JUNE, 1942 No. 3

THE JOURNAL OF APPLIED PSYCHOLOGY

Edited by
JAMES P. PORTER, Ohio University
Athens, Ohio

TABLE OF CONTENTS

DANIEL STARCH. The 100 Greatest Books Selected by 100 Qualified Persons 257

H. MELTZER. Mental Hygiene and Learning 268

EDWARD C. WEBSTER. A Follow-Up on Vocational Guidance 285

ROBERT B. SELOVER. A Study of the Sophomore Testing Program at the University of Minnesota. Part I 296

T. A. JACKSON, E. A. JEROME, AND N. SCHOENFELD. Experimental and Statistical Analysis of the Effectiveness of Deodorant Creams 308

ROBERT J. LEWINSKI. A Qualitative Analysis of the Kent Oral Emergency Test as a Clinical Instrument in the Examination of Naval Recruits 316

EDWIN E. GHISELLI. Estimating the Minimal Reliability of a Total Test from the Intercorrelations Among, and the Standard Deviation of, the Component Parts 332

WILLIAM R. TUBBS. A Study of the Interrelationships between the Adams-Lepley Personal Audit and the Bernreuter Personality Inventory 338

W. E. COOPER AND M. E. BONNEY. The Validity and Reliability of the Van Ripper Critical Angle Board in Diagnosing Handedness 352

C. W. Brown, L. B. Fisk, and H. P. Torkelson. The Effect of Changes in Vitamin A Content in the Diet upon Recovery from Glare Blindness 359

Alice I. Bryan and Walter H. Wilke. Audience Tendencies in Rating Public Speakers 371

L. Bald, F. K. Berrien, J. B. Price, and R. O. Sprague. Errors in Perceiving the Temporal Order of Auditory and Visual Stimuli 382

News and Notes 389

Book Reviews 393

New Books and Pamphlets 398

THE 100 GREATEST BOOKS SELECTED BY 100 QUALIFIED PERSONS

DANIEL STARCH

Daniel Starch and Staff, New York City

THE purpose of this inquiry was to determine what, in the judgment of competent persons, are the greatest books of all time. Numerous lists of "great" books, or "best" books, or books "everybody should read" have been prepared ever since there have been books. The Bible, itself, is not a single book but a collection of the most important religious books or scriptures selected by consensus of judgments of early church authorities from among works that had been written during a thousand years prior to about 100 A.D. Most lists have been prepared by individuals either alone or with the assistance of a few others for special purposes or from specific points of view—such as books we like, or books an educated person should read, or leading books published since a certain date, and the like.

Perhaps the two most important lists of great books assembled in recent times were the one compiled by the English naturalist and statesman, Sir John Lubbock, and the other begun by John Erskine, author and professor at Columbia University.

Sir John Lubbock published, in the *Pall Mall Gazette* on January 11, 1886, what he considered to be the 100 best books —books that were "necessary for a liberal education" or "most conducive to a healthy mind." He "excluded (1) works by living authors, (2) science, and (3) history, with a very few exceptions," which he mentioned rather in their literary aspects. The list was published also in the *Contemporary Review*. Various persons were then asked to comment

on this list and make suggestions. Comments and suggestions were received from James Bryce, John Ruskin, William E. Gladstone, Thomas Carlyle, Joseph Chamberlain, the then Prince of Wales who become King Edward VII, and others. Most of them approved Lubbock's list or suggested a few additions or omissions. Ruskin criticized it violently, crossing off about half of Lubbock's books including such important works as Aristotle's "Ethics," Marcus Aurelius' "Meditations," Confucius' "Analects," Smith's "Wealth of Nations," Sophocles' "Oedipus Rex," Euripedes' "Medea," Darwin's "Origin of Species," and Goethe's "Faust."

In the 1920's, Dr. John Erskine composed a list of sixty to seventy books as his choice of "the great works of Western culture." These books were the substance of a two-year reading course at Columbia University. Later, the list was modified and extended by members of the faculty of Columbia University and published by the American Library Association. The Erskine-Columbia list was further revised and extended by Professor Mortimer Adler at the University of Chicago and by the faculty of St. John's College at Annapolis where it became the core of its course of study. This resulted in a list of slightly more than 100 books and is published in Adler's *How to Read a Book* (1939).

So far as I am aware, no one has compiled a list of the world's great books by consensus of a large number of competent persons in accordance with a definite statistical procedure. Hence, I undertook, in 1937, to compile a list on such a basis. The procedure was as follows:

Four persons, broadly familiar with literature and the various fields of knowledge, made a list of all books which conceivably might be considered by competent persons to be counted among the 100 greatest books. No books published since 1900 were included. A generation or more must elapse before the permanent place of a book can be judged. This produced a total of 281 books. These were then arranged in alphabetical order and printed in mimeographed form. This

list of 281 books was then submitted successively to judges with the request:

> "Please make a check in front of each book that you would include among the first thirty-five great works and among these double check ten of the really greatest books. Please add at the end of the list the names of any others which you think should be included among the first thirty-five and then check them according to your opinion of their greatness. You are to regard greatness whatever you think it should mean. It should include, however, influence on mankind and civilization of an enduring character."

This procedure was continued until a total of one hundred judges had made their choices. The persons who were asked to make the selections included many distinguished authorities in the various fields of knowledge and literature. Approximately half of these one hundred persons were literary critics or professors of language and literature. The remaining half included persons distinguished in the various fields of knowledge — science, philosophy, economics, history, and government.

The returns were then compiled as follows: A weight of three and one-half was given to each book with a double check, and a weight of one to each book with a single check. The composite judgment yielded the following list of books with their respective scores:

THE 100 GREATEST BOOKS

Copyright, 1937, by Daniel Starch

Author	Title and Date	Consensus of Judgment Score
1. Numerous authors	The Bible—450 B.C.–200 A.D.	309
2. Shakespeare, W.	Hamlet—About 1600	211
3. Aristotle	Works—About 350–325 B.C.	198
4. Homer	Iliad—About 800 B.C.	197
5. Darwin, C.	Origin of Species—1859	187
6. Dante, A.	The Divine Comedy—About 1300	173
7. Plato	Republic—About 400–350 B.C.	166
8. Goethe, J. W.	Faust—1808	156
9. Confucius	Confucian Classics—About 525–480 B.C.	147
10. Milton, J.	Paradise Lost—1667	120

Author	Title and Date	Consensus of Judgment Score
11. Cervantes, M.	Don Quixote—1605	119
12. Mohammed	Koran—About 610–632	111
13. Chaucer, G.	Canterbury Tales—1386	97
14. Newton, I.	Principia Mathematica—1687	89
15. Kant, I.	Critique of Pure Reason—1781	88
16. Virgil	Aeneid—About 70–19 B.C.	84
17. Tolstoy, L. N.	War and Peace—1866	83
18. Marx, K.	Das Kapital—1867	80
19. Homer	Odyssey—About 600–500 B.C.	73
20. Bunyan, J.	Pilgrim's Progress—1678	72
21. Gibbon, E.	The Decline and Fall of the Roman Empire—1772	69
22. Bacon, F.	Novum Organum—1612	66
23. Sophocles	Oedipus Rex—About 468–406 B.C.	63
24. Buddha	Buddhist Suttas—About 480 B.C.	62
25. Smith, A.	The Wealth of Nations—1776	58
26. Aesop	Fables—About 570 B.C.	50
27. Montaigne, M.	Essays—1580	50
28. Hugo, V.	Les Miserables—1862	49
29. Plato	Phaedo—About 400–350 B.C.	49
30. Mendel, J. G.	Principles of Heredity—About 1850–1884	48
31. Burton Translation	Arabian Nights—1885	44
32. Shakespeare, W.	King Lear—About 1600	43
33. Aquinas, St. T.	Summa Theologiae—1265	42
34. Aeschylus	Prometheus Bound—About 500—456 B.C.	41
35. Hindu Sources	Bhagavad-Gita—About 200 A.D.	40
36. Calvin, J.	The Institutes of the Christian Religion—1536	40
37. Rousseau, J. J.	Social Contract—1762	39
38. Plutarch	Parallel Lives—About 100–120 A.D.	37
39. Blackstone, W.	Commentaries on the Laws of England—1765	36
40. Ancient Service Books	Book of Common Prayer—About 1500	36
41. Boswell, J.	The Life of Samuel Johnson—1791	34
42. Defoe, D.	Robinson Crusoe—1719	33
43. Shakespeare, W.	Macbeth—About 1600	33
44. Marcus Aurelius	Meditations—121–180	32
45. Emerson, R. W.	Essays—1841	32
46. Sophocles	Antigone—About 468–406 B.C.	32
47. Shakespeare, W.	Romeo and Juliet—About 1600	31
48. Thucydides	History of the Peloponnesian War—About 420 B.C.	31
49. Dickens, C.	David Copperfield—1850	30
50. Franklin, B.	Autobiography—1784	30
51. Locke, J.	Essay Concerning Human Understanding—1690	29
52. Bryce, J.	The American Commonwealth—1888	28

Author	Title and Date	Consensus of Judgment Score
53. Herodotus	History—About 490–480 B.C.	28
54. St. Augustine	Confessions—399	28
55. Whitman, W.	Leaves of Grass—1855	27
56. Caesar, J.	Commentaries on the Gallic Wars —About 50 B.C.	27
57. Stowe, H. B.	Uncle Tom's Cabin—1852	27
58. Lincoln, A.	Addresses—1858–1865	26
59. Malthus, T. R.	On the Principles of Population —1798	26
60. Galileo Galilei	Dialogue of the Two Chief Systems—1632	26
61. Omar Khayyam	The Rubaiyat (Fitzgerald)—1859	26
62. Rabelais, F.	Gargantua and Pantagruel—1533	25
63. Shelley, P. B.	Poems—About 1820	25
64. Machiavelli	The Prince—About 1512	24
65. Descartes, R.	Discourse on Method—1636	23
66. Fielding, H.	Tom Jones—1749	23
67. Hume, D.	An Enquiry Concerning Human Understanding—1748	23
68. Spinoza, B.	Ethics—1675	23
69. Thackeray, W. M.	Vanity Fair—1846	23
70. Voltaire	Candide—1759	23
71. Wordsworth, W.	Poems—1807	23
72. Rousseau, J. J.	Confessions—About 1770	23
73. Twain, M.	Tom Sawyer—1875	23
74. Dostoievski, F. M.	Crime and Punishment —1866	22
75. Epictetus	Discourses—About 90 A.D.	22
76. Harvey, W.	On the Motion of the Heart and Blood—1628	22
77. Euripides	Hippolytus—About 480–406 B.C.	21
78. Boccaccio, G.	The Decameron—1348	20
79. Kempis, T.	The Imitation of Christ—1471	20
80. Erasmus, D.	The Praise of Folly—About 1510	19
81. Swift, J.	Gulliver's Travels—1720	19
82. Shakespeare, W.	A Midsummer Night's Dream—1595	19
83. Faraday, M.	Scientific Papers—About 1833–1867	19
84. James, W.	Principles of Psychology—1890	19
85. Ibsen, H. J.	Doll's House—About 1880	18
86. German Sources	Nibelungenlied—About 1200	18
87. Shakespeare, W.	Merchant of Venice—1596	18
88. Pepys, S.	The Diary of Samuel Pepys—1660–1669	18
89. Aeschylus	Agamemnon Trilogy—About 500–456 B.C.	17
90. Bacon, F.	Advancement of Learning—1605	17
91. Burns, R.	Poems—About 1787	17
92. Macaulay, T. B.	The History of England—1848–1855	17
93. Keats, J.	Shorter Poems—1817	17

Author	Title and Date	Consensus of Judgment Score
94. Shakespeare, W.	Sonnets—1609	17
95. Dostoievski, F. M.	Brothers Karamazov—1880	16
96. Tennyson, A.	In Memoriam—About 1833–1850	16
97. Carroll, L.	Alice in Wonderland—1865	15
98. Cicero, M. T.	Orations—About 80–43 B.C.	15
99. Spencer, H.	First Principles—1862	15
100. Tolstoy, L. N.	Anna Karenina—1877	15

If the writings of Shakespeare are counted as one book the next six books to total one hundred are as follows:

101. Paine, T.	The Rights of Man	14
102. St. Augustine	The City of God	14
103. Veblen, T.	The Theory of the Leisure Class	14
104. Lucretius	On the Nature of Things	14
105. Schopenhauer, A.	The World as Will and Idea	14
106. Vedic Sources	Upanishads—About 500 B.C.	14

VALIDITY OF LIST

The validity of this list, I believe, is high, in the sense that another group of 100 competent persons at the present time would choose the same list with only minor differences among the lowest ten or fifteen books. As a matter of fact, the composite selection of the first fifty and of the second fifty of the 100 persons were practically identical. As a further corroboration, I asked one hundred librarians of large libraries, and also seventy-five recent college graduates engaged in business who had been out of college five to ten years, to make their selections. The results from both groups showed substantial agreement. Possibly a dozen books at the lower end of the list were different.

The final list as here presented may therefore be regarded as *the* one hundred great or greatest books, with some reservation as to the last ten or fifteen books. It has a reliability not possessed by any previous list for the following reasons: (1) It is the fruit of the combined judgment of a highly competent group of judges. (2) These judges included representative persons, recognized as authorities, from all main fields of knowledge. (3) The judgments were expressed and collected according to a specific statistical procedure.

HOW MANY GREAT BOOKS ARE THERE?

We can answer this question from our statistical data with some assurance. If we plot the scores for the one hundred books in decreasing order from left to right, as in Chart I, we note that the first group of books stands out far above the rest. The curve starts at a high point, drops sharply, and then pursues a slowly declining course. It appears, therefore, that there are not one hundred equally great or nearly equally

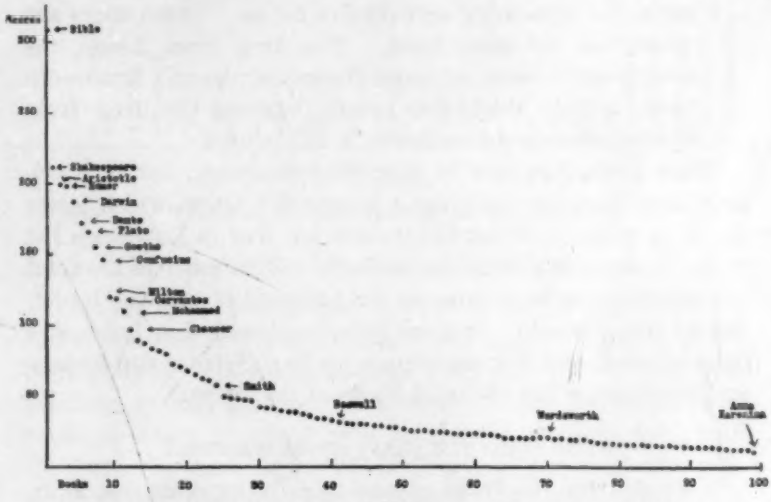


CHART I. Scores of the 100 books in order of decreasing size.

great books, nor seventy-five, nor fifty, but only about twenty-five greatest books. The chart indicates broadly that there are four groups of books:

First, the Bible. This stands out by itself as the book of books, with a score nearly 100 points or 50 per cent above the second book in the list, namely, the works of Shakespeare.

Second, the eight books, comprising the works of Shakespeare, Aristotle, Homer, Darwin, Dante, Plato, Goethe, and Confucius. These are, after the Bible, the really

towering monuments of literary and intellectual history which have had the most enduring influence on mankind and civilization.

Third, the sixteen books in the list from Milton to Adam Smith. These constitute the next group of great books. There is a drop of twenty-seven points from Confucius to Milton. Possibly this is not significant and these sixteen books should not be considered separately but in a continuous series with the eight in the preceding group.

Fourth, the remaining seventy-five books. These books are nearly on the same level. The drop from Aesop, the twenty-sixth book, to Anna Karenina, the one hundredth book, is only thirty-five points, whereas the drop from Shakespeare to Adam Smith is 161 points.

These groupings may be somewhat arbitrary, but it is evident that there are only about twenty-five works which tower above all others. These twenty-five are also on Lubbock's list or the Erskine-St. Johns' list, or both. Eighty-one of the total one hundred books are also on the Lubbock or the Erskine-St. Johns' list, or both. It must be remembered that Lubbock's list excluded, with few exceptions, books on science and history and the Erskine list excluded books of the Orient.

WHEN WERE THE GREAT BOOKS WRITTEN?

If we arrange the books chronologically by centuries, as in Chart II, we note the interesting fact that there have been two fertile periods in the world's history. The first period was the Greek-Roman-Early Christian era of about a thousand years ending around 200 A.D. The second is the renaissance era beginning about 1300 A.D. and continuing up to the present. Between these two eras there were a thousand years in which only five of the great books were produced. In fact, there were five hundred years between 600 and 1100 A.D. during which not a single work was created which was rated among the 100 great ones. Of the twenty-five greatest books at the top of the list, only one, the Koran, was produced during the

eleven centuries from 200 A.D. to 1300 A.D., whereas the millennium preceding 200 A.D. produced nine of the twenty-five greatest books. The Dark Ages obviously were dark from a literary point of view.

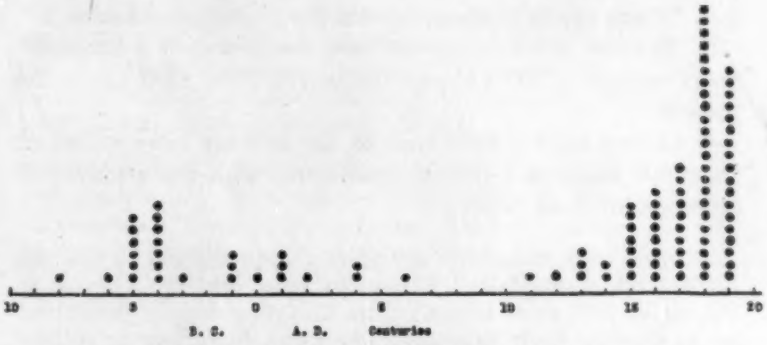


CHART II. Centuries in which each of the 100 books was written.

BOOKS READ BY COLLEGE GRADUATES

This list of 100 books was submitted to 142 seniors at the end of the year in the arts colleges at Stanford University, the University of Wisconsin, and the University of Florida, with the request that each one designate with the letter, R, those books of which they had read substantial portions and, with the letter F, those books which they had not read but with which they had considerable familiarity through other sources, such as reading or lectures.

The average number of the one hundred books read by these college seniors was twenty-four. The average number of additional books with which they indicated familiarity was sixteen, making a total of forty. Shakespeare's works had been read by the largest proportion, namely 96 per cent. The Bible had been read by 67 per cent with an additional 20 per cent expressing familiarity with it. At the bottom of the list were Newton's Principia Mathematica and Marcus Aurelius's Meditations, which had been read by, or were familiar to, less than 5 per cent of college seniors.

SUMMARY

1. A method has been employed which has yielded a list of 100 great books possessing a higher degree of validity than previous lists.

2. There are only about twenty-five "greatest" books.

3. The two productive eras were the period of a thousand years ending at 200 A.D. and the period from 1300 A.D to the present.

4. College seniors have read on the average twenty-four of these 100 books and possess familiarity with the contents of sixteen additional books.

NOTE: I wish to convey my cordial appreciation to the 100 persons whose combined judgments went into the making of the list of 100 great books. The following kindly permitted me to mention their names. (The word university or college is omitted from the names of institutions given in parenthesis):

Professors of English or Literature—Walter Barnes (New York), Philo Buck (Wisconsin), O. J. Campbell (Columbia), Hardin Craig (Stanford), Edward H. Gardner (formerly Wisconsin), W. H. F. Lamont (Rutgers), John Livingston Lowes (Harvard), R. L. Lyman (Chicago), Stuart Robertson (Temple), Howard J. Savage (Bryn Mawr), Homer A. Watt (New York), Karl Young (Yale).

Professors of other languages, ancient and modern—Barry Cerf (Reed), J. P. W. Crawford (Pennsylvania), H. C. Lancaster (Johns Hopkins), Kenneth McKenzie (Princeton), E. K. Rand (Harvard), David M. Robinson (Johns Hopkins), Albert Schinz (Pennsylvania), B. L. Ullman (Chicago).

Authors, literary editors and critics—Robert Barlow, Clyde Beck, Charlotte Becker, Grosvenor Blake, Van Wyck Brooks, Struthers Burt, Elrick B. Davis, Edwin F. Edgett, Max J. Herzberg, Chas. P. Johnson, Kent Knowlton, Sterling North, Paul Jordan Smith.

Journalists and professors of journalism—Frank Luther Mott (Iowa), Walter B. Pitkin (Columbia), William Allen White, Thomas F. Woodlock.

Professors of philosophy—John Dewey (Columbia), Will Durant (formerly California), G. T. W. Patrick (Iowa), Ralph Barton Perry (Harvard), Frank C. Sharp (Wisconsin), Edwin D. Starbuck (So. California).

Professors of history and government—Sidney B. Fay (Harvard), G. S. Ford (Minnesota), Roger Howson (Columbia), H. F. MacNair (Chicago), William B. Munro (Calif. Inst. of Tech.), Frederic A. Ogg (Wisconsin), Winfred T. Root (Iowa), Louis M. Sears (Purdue).

Professors of economics and sociology—Emory S. Bogardus (So. California), T. N. Carver (Harvard), F. Stuart Chapin (Minnesota), Charles F. Ellwood (Duke), Richard T. Ely (formerly Northwestern), Irving Fisher (Yale), Herbert A. Miller (Bryn Mawr), Robert E. Park (Chicago), E. A. Ross (Wisconsin).

Professors of various sciences—F. E. Bolton (Washington), W. W. Charters (Ohio State), Marian M. Cobb (New York Schools), Dexter S. Kimball (Cornell), Robert A. Millikan (Calif. Inst. of Tech.), Frank Schlesinger (Yale), Harlow Shapley (Harvard), T. M. Simpson (Florida), Lewis M. Terman (Stanford), E. L. Thorndike (Columbia), R. S. Woodworth (Columbia).

MENTAL HYGIENE AND LEARNING*

H. MELTZER

Psychological Service Center, St. Louis, Missouri

EXCOMMUNICATION VS. COMMUNICATION BETWEEN HUSBAND AND WIFE

ABOUT seven years ago in the city of St. Louis a thirty-seven-year-old, well-educated, professional man married a somewhat younger society woman. They both had very good backgrounds and were themselves decent, respectable, considerate people. After a few tussles about family approval, they were married and all went well—went well for many months; in fact, went relatively well until a child was born. And then there was complete discord about how the child should be managed. The wife tended to be on the reckless, careless side of management; the man on the more staid, conservative, over-protective side. The first time they were aware of the difference of opinion about child training, the husband noticed that the feelings of his wife were hurt and decided to remember never to mention that type of thing again. The wife, being equally considerate, began remembering the type of thing she said that seemed to hurt the feelings of the husband and never mentioned it, at least in his presence. More than a year went by and both the man and the woman remembered evermore to be evermore considerate by not mentioning the unmentionables that would hurt the feelings of the partner.

At the end of two years unmentionables filled the house atmosphere in the absence of outsiders. Two years of success-

* Presented before Sections I and Q panel on "Education and the Learning Process" at the American Association for the Advancement of Science meetings, Dallas, Texas, December 30, 1941.

ful excommunication resulted in emptiness and meaninglessness in the relationship of two well-meaning people. It was at this point that the husband called at eleven P.M. and, though not much of a cussing man, did some cussing and bewailing of his fate. Most of his language said in effect: "Damn it, I can't take it anymore. Here we are, two decent human beings and I think we still love each other, but there is nothing between us. We hardly say a word to each other." And he was correct. Even at the dinner table there was not much more said than, "Please pass this" and "Please pass that." Successful excommunication techniques had blown the life out of the home and the wind did not do it.

In this case it was a relatively easy matter to teach each how to communicate with the other a little bit more venturesomely, a little bit more courageously, more honestly, even at times more roughly and more dirtily. And the untouchable and unmentionable household became a place where more and more things were mentionable. Some vitality and meaningfulness returned in the home with the introduction and the maintenance of communication techniques rather than excommunication techniques.

There is little if anything in academic psychology that throws much light on learning processes in human relations of this nature. There is little about the development or persistence of such communication techniques considered in experimental or statistical studies of attitudes. Laws of learning based on experimental studies of animal learning will yield little about the learning processes involved in the relations between our husband and wife. The husband in this instance knew what educational psychology had to say about learning. And that was of no help to him and certainly did not keep him from making a substantial, though unwitting, contribution to the development of an emotionally sick relationship.

And this is how two decent and respectable people in the year 1934 had succeeded in living together in a way that made

both sick of each other and sick of life. In this case there was no need for a long-timed psychoanalysis. There was only need for re-evaluation of life meanings and the reconstruction of life attitudes that would make possible the emergence of meanings. It is an illustration of mental hygiene in a realistic manner, a consideration of human beings in action in real human surroundings, manifesting in one instance an attitude which made humanized learning impossible and, with increased understanding, manifesting attitudes which made possible the development of increased naturalness, expressiveness, and meaningfulness in life.[1]

PROBLEMS OF MENTAL HYGIENE AND LEARNING

The foregoing case is illustrative of case studies we wish to consider. They are cases of relatively normal people in relatively normal human relations. They are selected because they illustrate some facts about how mental hygiene can be acquired; how mental disease can at times be prevented; how emotional sickness can very easily be learned and reinforced; how some people in their efforts to be reserved and proper merely make themselves inaccessible not only to other people but sometimes to themselves; how feelings of understanding in human relations play a more important role in contributing understanding than intellectual understanding; how human motives of actual human beings are not only complex but are continuously susceptible in the dynamics of life to ambivalence, conflict, and at times tension; how learning processes play a role in the accessibility of all human relations, whether it is the acquisition of mental health or sickness; and how significant cultural and social factors, even facts of opinion, can be in helping direct a personality trend or a personality deviation.

[1] For a discussion of the learning process which does consider learning principles and conditions applicable to learning situations presented in this paper the reader is referred to a recent volume published for the Institute of Human Relations, *Social Imitation and Learning*, by Neal E. Miller and John Dollard.

The problems considered in this paper may be expressed in the form of questions, as follows: What are the barriers which have kept human beings from liberating their intelligence in their efforts to solve human problems? Why do many educated and intelligent individuals find it more difficult to overcome anxieties, phobias, and delusions than dull or uneducated people? How does it happen that a bright mother who, as she sees it, teaches her child obedience, respect, loyalty, love of art, and other select human virtues raises a child who lies, steals, feels love-starved, and is hostile, resistive, and suspicious? Why is it that such bright people can be so dumb? How does it come about that so many people searching for security, recognition, happiness, mental health, obtain results they are not seeking, and that so many people who mean to do so well by other people contribute toward the emotional sickness of those people? What is the nature of the difficulties and perplexities that they get themselves into? What can be done about it? What has been done about it in some instances? That is, how can personality realistically be considered, and what provisions for changing can be made? What are some limitations and possibilities in efforts to change human nature by clinical procedures?

WHAT ONE MOTHER TEACHES AND ONE CHILD LEARNS

James Harvey Robinson many years ago defined a college as a place where there is much teaching and no learning. In the next case I want to describe the home scene in which there was much teaching and also much learning but what the mother was teaching and what the child was learning did not resemble each other as far as attained goals were concerned. This is the case of a 14 year-old adopted girl who was expelled from a private school because she wrote love letters to a man who worked in the garden. She had a history of writing such letters for four years. Other problems her mother was concerned with were lying, cheating, and not doing well in school.

The mother was an executive, who married at the age of

thirty-eight. Home was an adult center which the mother administered. Mother even gave expert lectures on child study and did well in and out of town. She knew all the answers.

Father in the name of unification of purpose left everything to the wife in the raising of the child. Mother had a fear of pregnancy, a fear of having children, because her own mother died at childbirth. As you might guess, she has never made adequate sex and love adjustments with the husband. However, the husband never complained and had nothing but the best to say of her. Hence he said he loved his wife, adored her, respected her, and never did anything that would tend to hurt her feelings.

What did the child think of her parents, particularly the mother? In giving free associations concerning the mother, some responses she gave were: I can't tell if she loves me—Daddy thinks she is great but she gives me horrible punishments—she holds little things against me and talks against my real parents. Her reaction to father was: I love him and I think he loves me but he loves mother so much I can't tell him anything.

Now here is a mother, intelligent, dominant, successful on a large scale in her career and in community undertakings, teaching the virtues of humanity as she sees them—obedience, respect, loyalty, persistence, success, efficiency love of art, love of music, love of great people—at least, the people she considered great. This was what the mother tried to teach. Here is what the child learned: fear, suspicion, anxiety, emotional confusion, insecurity, repressed hostility, resistance to authority, and emotional ambivalence.

There are many educational ramifications of this one case study. Among the list of concepts illustrated are: the effect of conflicting purposes in personality development—the mother's notions of the child's needs and the child's needs as perceived by the child; the effect of intellectual virtues of the mother who meant well on a child who felt love-starved; an-

other illustration of the projection of ambition by the mother on the child; an illustration where emphasis is on compulsion and duty rather than on contribution and training; how lack of feelings of belongingness can influence phantasy thinking and distracted attention in school; how a dominant mother prevents a child from building up open relations with the father she loves for fear that she would not get a hearing from him because of his love for mother. And it is again a lesson in human modes of expression being repressed and excommunication techniques developing. Such efforts as were made for understanding were intellectually toned. With them did not go an emphasis on the feeling of understanding. Even when the mother's impressions of the child were accurate, they did not serve to influence the child in giving her a feeling of being understood by the mother. A distrust and a suspicion attitude tied up with ambivalence were practically always present.

This case also might be given as evidence for the fact that the Norwegian psychologist, Jaederholm, used to emphasize; namely, that neither the one hundred per cent efficient individual nor the one hundred per cent pasteurized individual can be considered a worthwhile individual, at least not a healthy-minded individual. In Jaederholm's words,[2] "A reservation should be made against the idea, that the last goal of human personality should be a complete adjustment to his social life. The perfectly socialized being is to my mind a very unhappy ideal which leads, when realized, to far too much dependence on others, and on their very inadequate judgments of sympathy and antipathy. In such a state, the higher values of life may, if not entirely overlooked, be pushed very much into the background. Hence the shaping of life according to one's own best insight and aspirations becomes practically impossible. It may even be assumed, and with no great error, that a complete socialization of the individual (though per-

[2] See *Behavior Levels and Mental Hygiene*, by Gustav A. Jaederholm, Department of Public Welfare, Illinois, Series B, No. 102.

haps with some individuals it may lead to a greater common-sense-happiness), will be shown up in the end as an arch-enemy of some of the noblest powers in man's personality! It may destroy his legitimate desire to make out of his personality, with all its weaknesses, the best and most valuable thing possible."

To the orthopsychiatric group this case would be another illustration of the fact that the feeling of security and belongingness that a child may get in the early period is more important in the development of a wholesome, pleasantly assertive personality than all the lectures that follow thereafter. Or, to use the words of Dr. Plant, it illustrates realistically and well the fact that of the two dimensions of personality he emphasizes—the "who am I" and the "what am I"—the "who" is more significant. By "who" is meant the feeling of belongingness, the feeling of being wanted, regardless of how good or bad a person is. By "what" is meant achievement. The implication, of course, is that the dynamics of the learning process in alive human beings which result in achievement are determined to an extremely large extent by the early signs of the kind of feeling of acceptance or rejection, the presence or the lack of it, that the human infant gets in its early years.

FROM COLDNESS AND RIGIDITY TO WARMTH AND FIRMNESS

A nine-year-old boy from a relatively high middle class was referred by his mother. The reasons for referral were many. They included such complaints as the following: The child lacks control—he hits and spits. He is high strung. He has extremely bad dining-room manners and he is erratic. He is over-active. He is screamy. He clings too much to his mother.

An examination of the child revealed him to be a very bright child (121 IQ) who had an extremely effective and insightful way of learning. He was a repressed child, would answer only when questioned, and rarely had a happy expression on his face. An interview with the child revealed the meaning of the symptoms the mother complained of; namely, they were all the

effects, as far as the child as a learner was concerned, of mother's efforts to do practically all of her teaching of the child at home as emergency teaching in urgent situations. As far as the child was concerned, life with his parents was surrounded by the nagginess of a mother on one side and a sour-puss father on the other. Father, as perceived by the child, would work in duty-bound fashion in the shop in the cellar with him, but never took a pleasant expression with him, and always in response to the child's efforts expressed disappointment with him. This in spite of the fact that the boy had mechanical abilities far above his age level and succeeded in constructing things that were deserving of praise for a boy of eleven or twelve, though he was only nine.

In this particular case, the child was not seen again; the work was done with the parents. In the first session after the examination, the parents were given a lesson in education for naturalness and spontaneity. The meaning of the symptoms in relationship to the child's conception of the mother's nagging and the father's expressions of disappointment were gone over thoroughly and also laughingly.

It may be of interest at this point to describe what happened in the first interview. Knowing of the parental stiffness that existed in both parents, the clinician ventured to call both of them on their first visit after the examination by their first names. Mother looked at father; father looked at mother. Shock and amazement and some unpleasantness registered in their gestures though their facial expressions registered little that could be described. Expecting and observing this, the clinician then asked them if they really meant to help the child. Both of them, of course, said definitely that was their interest. After many repetitions of this, they were then advised there was no sense in considering the findings of the examination until both of them permitted themselves and gave themselves license to be somewhat more human, more natural, more easy, more spontaneous, with some efforts of glad willingness to learn. They were advised that calling them by their first

names was one way of breaking through their stiffness. Some more looks and then father, in this case, let out a good laugh, and that was the end of his resistance. The mother still showed some reserve at first but got into the spirit of the thing in a few more minutes, and thereafter a pretty thorough consideration with an emphasis on those things in the history that needed emphasizing was possible.

Four days of new attitudes manifested by both parents and the mother called and came in for an appointment. With glad sadness she cried a bit but warmly said, "What I have to report I suppose will amuse you. I liked the results but it pains me." She then quoted the child as saying before going to sleep the night before: "Mother, do you really love me as much as you have been telling me the last few days, and does Daddy love me that much? I am so happy."

Since that time the family has organized a weekly family session, and balance as well as spontaneity are being provided for in the training processes in that home. What was once a proper and rigid home is now a natural and expressive home in the making.

THE ROLE OF CULTURE CONFLICTS ON PERSONALITY

In recent years educators along with sociologists have done much talking about the effect of culture on personality. But most of their illustrations have been taken from culture conflicts of first generation families in America, as illustrated by Thomas' *Polish Peasant in Europe and America.* We propose to present a case to demonstrate that one need not do much searching outside of the United States to see the effects of culture conflict. The case selected is of a 21-year-old boy who was raised in the city of St. Louis under very proper hard-shelled Baptist auspices. While he was in St. Louis, raised in the same neighborhood with boys who went to the same church and the same school, all was well. And then one year he decided, with parental approval, to take an engineering course at a university located in Indiana. In St. Louis

he was raised on a diet of three taboos—no smoking, no drinking, no dating. In the Indiana university in one year he learned to do all three evils. The second year he went to Washington University in St. Louis, his home town, and did not return to the university in the state of Indiana. His grades at Washington University did not indicate anything related to interest, effort, or learning. He was flunking practically all courses and was getting himself into serious trouble with the faculty.

On the first interview at the psychologist's office, he asked permission to smoke. He took out a cigarette, took one or two puffs at it, and with a shaky hand put it out; again asked permission to smoke, again two puffs, again used a shaky hand to put it out. And here was the story.

On his return from the Indiana school he could not study in his home because he would be feeling his pockets for cigarette packs fearing that his dad or mother might search and find them. He washed out his mouth to erase the smell of the drinks from the night before or the cigarettes of an hour before. At the dinner table he fidgeted at his pockets and his mouth to cover up his guilty feelings concerning his sinning. When in his study room in his own home, the sound of steps, even in hours when he knew his father was not home, would make him restless and fidget at his mouth and wash his mouth. Such studying as he did succeed in doing certainly did not add to his knowledge of things and did not show in his school work.

In this case we received the boy's permission to interview the father, and it was possible to re-educate the father so that he, at first against his will, agreed to let the boy do all the three evils he had forbidden him and gave him this license in the presence of his dean. Some more interviews, some more open attitudes, and studying began to register and grades began to improve.

This is not only an illustration of culture conflict between the old and the young generation, between mores in one city

and in another, one campus and another perhaps, but also is another illustration of how excommunication techniques can develop and dehumanize the relationship between a father and a son and how open communication can humanize a relationship. It also illustrates the complexity as well as the possibility for conflict in human needs and wants. It is easy for psychologists to talk about and list urges and drives and desires or for sociologists to list wishes, but in actual life situations the needs are all too often in conflict and much more complex than is indicated or implied in lists discovered in books. Much more realistic is the picture given by Lynd in his book, *Knowledge for What,* in which his emphasis is on the fact that culture encourages ambivalence. His illustrations are the ambivalence between saving and spending; between playing safe and "nothing ventured, nothing gained"; between "you've got to look like money to make money" and "spend for what you want"; between brains and charm; between right in theory and wrong in practice; between loyalty and looking out for number one; between being efficient and being human.

PUBLIC OPINION AND MENTAL HEALTH PROBLEMS

In many cases the factor of opinion can be shown to be more potent in influencing behavior and determining an attitude than facts. Since the facts of mental health and disease are not so well known, many opinions have been released as facts. That the style of communication in the emotional atmosphere created in the home helps characterize the behavior of parents as well as children in the home is observed by many people. Not sufficiently realized is the fact that when people deal with abnormalities in behavior or extremes in behavior—say, mental disease or even feeblemindedness, the attitudes of the people who have to deal with feeblemindedness are very often more important than the accuracy of the diagnosis of feeblemindedness.

In one case, for example, a mother had three children. One,

the youngest, was an idiot, and the mother spent the better part of twenty years focusing all her attention taking care of the idiot. During that period of time, the child was committed to an institution twice but taken out almost immediately. In the emotional atmosphere created by the mother paying almost full attention to the idiot and no attention to speak of to the other children, two youngsters, one of very superior intelligence and one above average, have been unintentionally directed into schizophrenic trends. Here are two socially worthwhile human beings not only disregarded but distorted by a mother because our provisions for feeblemindedness are in the form of institutions where we can place a child and in the form of clinics where a child can be examined. The most pivotal factor in this case, the re-education of a mother to face the facts and to act accordingly, was completely neglected by a clinic. In three sessions twenty years later she was educated to do that and the child has been in an institution the last many years. Obviously this has helped to some small degree the two more promising sons. Also obvious is it that if it had been done a long time before they would have been helped much more significantly and much more promisingly.

Even in this case, and much more so in the others cited, the problem is not in the diagnosis, after which treatment follows, but the problem is in the dynamics of understanding, fuller and fuller understanding, expressive understanding, in the light of which the realities can be managed.

In an education for reality it may be well to keep in mind the teaching of John Dewey concerning man, nature, and society (*Human Nature and Conduct*): "When man had no scientific knowledge of physical nature, they either passively submitted to it or sought to control it magically. What cannot be understood cannot be managed intelligently. The intelligent acknowledgment of the continuity of nature, man, and society will alone secure a growth of morals which will be serious without being fanatical, aspiring without sentimentality, adapted to reality without conventionality, sensible

without taking the form of calculation of profits, idealistic without being romantic."

EDUCATION FOR INTERDEPENDENCE VS. OMNIPOTENCE

All the foregoing case studies clearly illustrate that the problem in all instances is not only one of style of communication but that the style of communication is related to the level of emotional maturity of the participating individual. In the limited time we cannot discuss in detail the processes involved in growth and maturity but we can glance at certain phases of genetic processes; namely, the growth from fictitious independence to interdependence, that is illustrated in all our case studies. From the point of view of social significance, the first stage of growth might be considered the stage of fictitious independence. This is followed by a long period of dependence, first on one person and then on more and more persons in an expanding world. And with increasing growth and confidence and discrimination comes the more matured attitude of interdependence. Growth, then, goes from fictitious independence by way of a long period of dependence toward interdependence.

The period I am referring to as fictitious independence is often referred to in psychoanalytical literature as the period of omnipotence. The implication is that the child is lord of all he surveys. His desires, few in nature, are not only satisfied for him without any conflict but are even anticipated by the mother. When desires are so easily satisfied without any realization of limitations, omnipotence prevails. Such omnipotence in the period of infantile growth is natural and normal. Growing up, however, implies outgrowing earlier stages and growing into new stages more in accord with possibilities for maturing. The first time the child sets out, because of increased physical growth, to do something on his own, such as taking his first steps alone, and gets some distance away from mother and loses his security, he is immediately confronted by the meaning of conflict in growing up, which is experienced

throughout life; namely, this: that the person asks himself, in a sense, "Should I stick closer to my earlier form of satisfying desires without making an effort and feel secure in a small world, or shall I try a life in a somewhat larger world but at the loss of comfort and at the price of taking a risk and making an effort?" And making a decision of this kind, whether to explore or to retreat by way of techniques that in literature are often referred to as defense reactions, evasive reactions or tension-making reactions, is an old story.

In this initial experience, the child with its first fall becomes aware for the first time of the fiction of his earlier omnipotence and the reality of his own limitations. If the child is guided sensibly, he discovers that he can well afford to live in a larger world and take an exploring attitude toward it. If he becomes aware of his littleness but, instead of getting a low feeling from it, feels good about it, since admitting his own dependence on mother at the outset often gives more freedom to mother's movements as well as his own, he can gain for himself an enriched security in a larger sphere of influence.

Of course, the foregoing attitude of the child toward his growing self in the earliest phases of his life is entirely interpretative and is not the result of the child's reporting all this. It is based more on adults' memories of their childhood life than it is on the actual reporting of very young children.

After the child makes his first discovery that he has limitations, then he transfers his omnipotence toward mother, who becomes the absolute authority. In due time, however, the child discovers that mother also has limitations and has to depend on still others. And so on for his growing awareness of limitations and assets in all human beings. When a human being has this realization and has a glad willingness to act on it, he is then ready for the more matured stage of interdependent behavior. But in the world today, quite obviously many individuals have apparently stopped growing somewhere along the line of this genetic picture. Hostility and suspicion are the end-products of repressed love all too often. There

are also symptoms of pretensive self-sufficiency, of unadmitted limitations, of being either overwhelmed by omnipotence or, by the use of the omnipotence drive, overwhelming others. When life is circumscribed, emotional richness has no chance for growth. Guilt and rage become the outer dimensions of emotional life. Guilt becomes associated with feelings of shame and the need for punishment; range means a projection of that shame and blame on others.

EDUCATION IN CAPTIVITY VS. EDUCATION FOR INTERDEPENDENCE

In the light of the foregoing facts of genetic growth, it becomes obviously important to consider in training processes the significance of emotional acceptance on the part of the learner at the outset. Otherwise we have the calamity that has previously been referred to in this paper and also illustrated; namely, that without emotional acceptance, teaching and learning do not resemble each other. What many teachers teach and what students learn are quite often not at all the same. Those who are at all familiar with animal psychology remember that Thorndike's monkeys could show nothing but trial and error learning when not given an opportunity to learn anything else in puzzle boxes. But when Kohler, the Gestalt psychologist, gave his monkeys the freedom of the yard and educated them not in captivity but in free space, they manifested the ability to perceive relations or, as he called it, insight. Certainly much of the education that goes on in our school systems is in the nature of education in captivity. And the implications of the facts of emotional growth are that no learning worthy of the name takes place in such stage settings.

Obviously it is important, since all of us have limitations, that when confronted with limitations in that stage of growth that might be called the fumbling area of selective dependence, we have attitudes that carry us on to interdependence rather than regression to omnipotence. Regression to omnipotence means keeping oneself to oneself, not having the desire to give and take with another human being, the adoption of a style

of communication technique where only a part of a person responds to part of another person but no whole person responds to another whole person. It means the growth of inaccessibility, the development of emotional sickness, whether it be colored with false pride or fear, or both. Interdependence implies the possibility for development of integrity, the possibility for development of mutual understanding with another human being, the possibility for perceiving insights in the light of all the intelligence within the individual without blind spots and without self-pity.

CONCLUSIONS

From the foregoing considerations of the case studies presented, a number of tentative generalizations emerge. For purposes of brevity let us express them in outline form.

1. Mental health can be acquired and so can emotional sickness.

2. Mental hygiene is concerned with the learning and living of actual human beings in all their human surroundings and never with sound body in a sound mind entities.

3. The learning processes for the acquisition of mental health and the acquisition of emotional sickness are quite similar in nature; namely, trial and error processes are manifested in both, occasional inductive leaping, and configurational learning.

4. The attitudes that make for mental health, however, are distinctly different than those which make for emotional sickness. The learning attitude which makes for mental health helps develop techniques of communication, which help the individual become more accessible to himself and thus more understandable to others as well as himself. Characteristically, an attitude that makes for emotional sickness is one which helps the individual become less accessible to himself and thus less understandable to himself and others.

5. Without emotional acceptance on the part of the learner, what the teacher teaches and what the learner learns do not yield the same results or goals.

6. Pivotal in learning to acquire mental health rather than

emotional sickness is the shift in the outlook of the patient or learner when he begins to talk and think and feel less about a condition and more about learning to live.

7. In the realities of human learning, as illustrated in clinical studies, ambivalence, conflict, and complexity are more apparent than in experimental studies of learning reported in psychology books.

8. Without a firm foundation obtained from an early experience in feeling wanted, in acquiring a feeling of belongingness, achievement is likely to be relatively discontinuous and spurtive.

9. Functional firmness rather than rigidity or looseness is desirable in personality organization.

10. It is highly worthwhile to emphasize the need for education for emotional maturity in all learning, in home, school, and other learning situations.

11. Without the acquisition of a sense of selective dependence which can grow into a more matured attitude of interdependence, the regression to omnipotence is all too easily acquired by human beings and results in Hitlerism and its many cousins.

12. There are many modes of human communication. Of all of them the use of language is the most human, but the use made of language by some people serves to excommunicate them from their fellows rather than to help them understand their fellows or to be understood by their fellows.

13. Finally, the kind of peace that the mental hygienist would advocate is not the kind that passeth all understanding. This all too often goes with excommunication techniques and pretensive understandings which explode into misunderstandings. In industrial relations, family relations, and school relations, as well as international relations of the Munich Pact type, there is no promise in the peace that is a fiction, a pretense, an hypocrisy, a peace that passeth all understanding. There is only one kind of peace the mental hygienist would advocate and that is based on the kind of understanding that makes further understanding possible.

A FOLLOW-UP ON VOCATIONAL GUIDANCE

EDWARD C. WEBSTER

The Psychological Institute, Montreal

FROM June 1934 to December 1937 the writer gave psychological examinations to 125 individuals whose problems were such that the cases could be described as vocational or educational guidance. In December 1939 a letter enclosing a follow-up questionnaire was mailed to each of the 125 clients. A second letter was mailed in April 1940 to parents and examinees who ignored the first appeal. A total of sixty-six replies was secured. Telephone calls and personal contacts provided information concerning present occupations and academic achievement of an additional 15 individuals. Thus, data are available in reference to 81 of 125 vocational and educational guidance cases examined from 2 to 5 years prior to the follow-up.

A DESCRIPTION OF THE PSYCHOLOGICAL EXAMINATIONS

Certain features are common to all the psychological examinations subjected to scrutiny in this investigation.

(1) Each examinee was given both performance and paper and pencil aptitude tests determined by the particular problem.

(2) Each completed at least one personality inventory and a number of prepared questionnaires dealing with habits, interests, and attitudes.

(3) The parents, in most instances, also completed a detailed questionnaire in reference to the examinee.

(4) Testing of each individual required from two to four hours. The principal examinations employed were: Otis S.A. Tests of Mental Ability; N.I.I.P. Clerical Test, the Minnesota

Paper Form Board and Assembly Test; a measure of introversion and one of sensitivity of feelings prepared by the writer. These were supplemented in particular instances by a wide range of examinations no one of which was used a sufficient number of times to warrant special mention.

(5) Test and questionnaire results provided a frame work around which interviews were developed. In two-thirds of the cases, from one to three hours were devoted to interviewing; in the remainder of the examinations from 5 to 10 hours were spent on the interview.

(6) A written report averaging between five and six pages was prepared on completion of the examination. This gave an analysis of the individual based upon test results and the interviews as well as recommendations of an educational or vocational nature.

This report concluded examinations in 1934. Those in subsequent years, however, included two later interviews—one with parents, the other with the examinee—at which the report was discussed.

DESCRIPTION OF THE FOLLOW-UP CASES

The 81 cases (75 boys and 6 girls) for which follow-up data are available, were examined between June 1934 and December 1937. Actually 65.5 per cent were studied in 1934 and 1935 as shown in the following figures:

Year of examination	1934	1935	1936	1937
Minimum elapsed time between examination and follow-up	5 years	4 years	3 years	2 years
Per cent of cases	55.6	9.9	18.5	16.0

The high proportion of 1934 cases results from the fact that in this year a free vocational guidance service was provided to 72 boys graduating from high school. Follow-up data were secured for 37 of these. All other cases reported were private paid examinations for which the fees ranged from $15 to $70.[1]

[1] There was only one difference between paid and free examinations: in the latter, group tests and questionnaires were completed by 8 to 10 boys at a time; in the former each case was dealt with individually.

Individuals ranged in age from 12 to 42, although 80 per cent were between 16 and 20. Two were over 30 and 6 were younger than 16.

The educational background of examinees is shown in the following figures:

	No. of cases
Some elementary school training	1
Elementary school graduates	4
Some high school training	10
High school graduates	60
Some university training	5
University graduates	1

All but eleven (13.5%) were students or had completed a course of study within two months of the examination.

Only one individual in the group had an IQ which is significantly below average. As a whole, the group was superior to an unselected sampling of the population on aptitude tests.

CRITERIA OF OCCUPATIONAL AND EDUCATIONAL ADJUSTMENTS

It is extremely difficult to obtain an adequate criterion for occupational adjustment. Obviously, neither permanency of work nor income provide a fair basis for judgment. The follow-up questionnaire sent to examinees was designed to provide information upon educational and vocational history and adjustment subsequent to the examination.

Occupational Adjustment:

The following specific questions were asked concerning work history:

> At what occupation is the examinee now engaged?
> Has this occupation been found satisfactory?
> In what ways, if any, is it unsatisfactory?
> What other occupations has the examinee held?
> Which of these, if any, were selected, partially at least, as a result of the vocational guidance examination?
> Why were these latter positions given up?

Answers to these questions together with some fifty letters and oral communications or notes written on the questionnaire forms provided considerable data upon which subjective estimates of vocational success could be based.

All available work history data were considered for each examinee and he was rated as "satisfactorily placed," "unsatisfactorily placed," or "undetermined," in his present job. The vocational guidance report of each examinee was re-read and a note made of the positive and negative recommendations. These were then compared with the actual work history. Table 1 was prepared on the basis of this comparison.

It should be noted that if the examinee entered an occupation recommended by the psychologist and left this career because of personal dissatisfaction or inability to perform his duties, the prediction for that case was judged "incorrect" regardless of other factors.

In a number of instances the examinee had had no occupational experience at the time of the follow-up. These cases were discarded.

Educational Adjustment:

Examination results provide a ready criterion for educational success. If a young man followed a course—whether a six month secretarial or a seven year Arts and Law—without failure he was rated as "educationally successful." If he attempted the course and did not complete it because of general failure he was rated "educationally unsuccessful." Between these two extremes are a number of steps indicative of relative success or failure. Each examinee's educational history was studied from replies to the following questions:

What academic work, if any, was followed subsequent to the psychological examination?
Where was it taken?
What degree or other recognition was secured?
What failures (general or in particular subjects) occurred?

Were educational courses selected, at least in part, on the basis of the vocational guidance examination? (check answer) Yes , No , Uncertain .

Each guidance report was re-read and a note made of all positive and negative educational recommendations. Table 2 was then compiled from a comparison of these two summaries.

Cases were discarded if no educational recommendations were made during the examination or if the examinee proceeded no further in his educational career.

ACCURACY OF PREDICTION

Vocational:

Fifty-five vocational recommendations were made where the examinee had work experience subsequent to the examination. Table 1 shows the accuracy of these predictions.

The total correct predictions are the sum of those where recommendations were acted upon and found satisfactory plus cases where the individual failed to succeed when he acted in opposition to recommendations.

TABLE 1

Accuracy of Vocational Predictions

Accuracy	No. of cases	Percentage
Recommendation acted upon and found satisfactory	28	51.0
Failure to succeed when acting in opposition to recommendation	8	14.5
Recommendations acted upon and found unsatisfactory	1	1.8
Success when acting in opposition to recommendations	8	14.5
Impossible to judge whether prediction is correct	10	18.2

The total failure in prediction includes cases where the examinee acted in accordance with the recommendations and found them unsatisfactory and instances where he succeeded although acting in opposition to the recommendations.

In 65.5 per cent of cases the predictions were proved correct and in 16.3 per cent they were proved false. Ten cases were

impossible to judge either because the examinee entered an occupation not considered at the examination, or positions have not been held for a sufficient length of time to warrant a judgment.

In evaluating the vocational recommendations, these doubtful cases may be discarded as the sample is small and they distort the percentage figures. If this is done, it will be seen that accurate predictions were made in 80 per cent of cases which have progressed to the point where definite decisions can be reached.

Educational:

Sixty-one educational recommendations were made where the examinee followed a longer or shorter course of study subsequent to the examination. The accuracy of prediction in these instances is shown in Table 2.

TABLE 2

Accuracy of Educational Predictions

Accuracy	No. of cases	Percentage
Academic success in particular courses accurately predicted	37	60.6
Academic failure in particular courses accurately predicted	14	23.0
Academic prediction proved incorrect	4	6.6
Impossible to judge accuracy of prediction	6	9.8

The educational future of the examinee is successfully predicted if he succeeds in a course recommended or if he fails in a course taken in opposition to the recommendations. Similarly, the predictions are incorrect if the examinee fails in a course of study recommended or succeeds in one opposed by the psychologist.

Accurate prediction of educational success or failure was made in 83.6 per cent of the 61 cases. Errors in prediction were limited to 6.6 per cent of the recommendations while in

9.8 per cent of cases where recommendations were made, study was continued in directions not considered. Hence, in these six instances it was impossible to judge the accuracy of prediction.

These doubtful cases may be discarded from further consideration inasmuch as they have no direct bearing upon the accuracy of prediction. When this step is taken, it is evident that 92 per cent of educational predictions proved correct.

Combined Educational and Vocational Predictions:

There are 116 instances where vocational or educational recommendations were made and where the examinee has a work or academic record subsequent to the examination.

Table 3 has been prepared from Tables 1 and 2 to indicate the accuracy of combined predictions.

TABLE 3

Accuracy of Combined Educational and Vocational Predictions

Accuracy	No. of cases	Percentage
Correct predictions	87	75.0
Incorrect predictions	13	11.2
Doubtful cases	16	13.8

Definitely correct predictions were made in 75 per cent of these cases and incorrect predictions in 11.2 per cent. When doubtful cases are discarded, the percentage of accurate predictions is raised to 87.

SUBJECTIVE APPRAISAL OF GUIDANCE

Accuracy in prediction is the essential criterion of the value of guidance. But, if guidance is to have practical value, it must be approved by the persons guided. Consequently the following question was asked: "In view of all events which have occurred since the examination, do you believe the vocational guidance was (or would have been had recommendations been followed): (Check appropriate answer) extremely valua-

ble ( ), of definite value ( ), of some slight value ( ), of dubious value ( ), a waste of time ( ), definitely harmful ( )."

Fifty-eight of the 66 persons returning the formal questionnaire answered this question and their replies are presented in Table 4.

TABLE 4

Subjective Evaluation of Guidance Examination

Opinion	No. of cases	Percentage
Extremely valuable	5	8.6
Of definite value	28	48.3
Of some slight value	15	25.8
Of dubious value	6	10.3
A waste of time	3	5.2
Definitely harmful	1	1.7

Eighty-two and seven-tenths per cent of cases report that the psychological examinations were valuable. This figure includes 56.9 per cent of replies which stated that the examinations were of definite value or extremely valuable. This subjective evaluation of the examination is considered very satisfactory in view of the fact that many recommendations were not as optimistic as parents and examinees had anticipated.

Available evidence indicates that this approval of vocational guidance does not result solely from the fact that the educational and vocational recommendations were helpful. Thus, it is observed that:

(1) Sixty-three replies were secured to the question "Was the occupation (now held) selected, at least in part, on the basis of the vocational guidance examination?" Thirty-three per cent of the answers were "Yes," 54 per cent were "No," and 13 per cent were "Uncertain."

(2) A similar question was asked respecting educational courses taken subsequent to the examination. Replies were secured in 57 instances. The examination was given at least

partial credit for the educational decision in 52 per cent of these cases. In 32 per cent of the 57 replies credit was withheld from the psychologist, and the remaining 16 per cent were uncertain.

Obviously if the judgment concerning the value of the examination was based solely on the credit given the vocational and educational recommendations, less than 82 per cent of persons would report the examinations of value. Examination of replies in reference to all available information leads the writer to the generalization that the psychologist is credited in assisting in the final vocational or educational decision provided his recommendation is not an approval of a course of action already tentatively decided upon by the examinee.

The fact that examinations had value apart from specific vocational and educational recommendations is confirmed by letters and notes which accompanied fifty of the returns. An improved social and emotional adjustment as a result of the examinations was reported in many instances. Others stated that they could approach their own problems more intelligently as a result of the vocational guidance. Such benefits account for the discrepancy between the subjective evaluation of the whole examination and the appraisal of specific vocational and educational recommendations.

Further indirect evidence of the examinee's appraisal of vocational guidance was obtained through a number of supplementary questions. Six general statements were presented in the follow-up questionnaire. Respondents were requested to indicate agreement or disagreement. The statements together with the results obtained from 55 individuals are presented in Table 5.

These 55 individuals had personal experience with psychological vocational guidance. Apart altogether from the one case in which they were interested, what do they think of guidance? Their opinion will undoubtedly have been affected by their own experience. Vocational guidance as a psychological technique will succeed or fail depending upon the support given it by persons who have received advice.

TABLE 5

Beliefs of Parents in Reference to Guidance

Statement	Agree	Disagree	Doubtful
	%	%	%
Vocational guidance should be given all adolescents when they leave school	76	2	22
Vocational guidance should be given problem students	94	0	6
Vocational guidance is of no practical value	4	90	6
By-products of the examination, *e.g.*, recommendations dealing with personality, social adjustment, etc., are of real value	87	2	11
Guidance provides a good substitute for the current practice of placing an adolescent in a school or job to "see if he can make good"	80	2	16
Guidance should be regarded as a luxury rather than a necessity	13	78	9

Evidently guidance is considered valuable by these clients. There is almost complete agreement that it should be given students who are problem cases; only 13 per cent believe it is a luxury; about 90 per cent believe that the by-products of the examination have real value; and four out of five consider it a good substitute for the older "try and try again" method.

While a technical evaluation may indicate the possibilities of vocational and educational guidance, only the public which has been subjected to these techniques can give final practical approval to the psychologist's work. It is obvious that this approval is given by the great majority of clients traced in this follow-up study.

SUMMARY AND CONCLUSIONS

1. A follow-up questionnaire study was made by mail in 1940 of 125 clients given vocational and educational guidance between 1934 and 1937. Data upon which this report is based were secured from 81 persons.

2. Fifty-five vocational recommendations were made where the examinee had a subsequent work history. Correct predictions occurred in 65.5 per cent of cases, incorrect ones in 16.3 per cent, and no decision could be reached in 18.2 per cent.

3. When consideration was given only to cases which had progressed to the point that definite comparisons could be made, the predictions were accurate in 80 per cent of cases.

4. Sixty-one educational recommendations were made where the examinee had a subsequent educational history. Accurate predictions occurred in 83.6 per cent of these cases. Errors were limited to 6.6 per cent. The remaining 9.8 per cent followed courses of study not considered at the time of the examination.

5. Ninety-two per cent of educational predictions proved correct when cases are discarded where the examinee followed courses not considered at the time of the examination.

6. Combining educational and vocational recommendations for all cases, 75.0 per cent of predictions were correct, 11.2 per cent were wrong, and in 13.8 per cent of cases comparisons could not be made.

7. When doubtful cases were discarded, 87 per cent of 100 predictions were found to be accurate.

8. Some 82 per cent of clients reporting, stated that the examination had been found of value. Furthermore, the examinees definitely believe in the practical value of vocational guidance. Nevertheless, only 52 per cent admit that the educational recommendations were helpful and only 33 per cent found the vocational recommendations of value. The supporting data suggest that the psychologist is given credit, not only for tangible help in the choice of educational courses and careers, but for more important help in correcting personality deficiencies. So far as actual vocational recommendations are concerned, the psychologist must, to receive specific credit, suggest careers not contemplated by the examinee.

A STUDY OF THE SOPHOMORE TESTING PROGRAM AT THE UNIVERSITY OF MINNESOTA—PART I*

ROBERT B. SELOVER

The Prudential Insurance Company of America

THE increasing emphasis on the need for a "liberal education" as a preliminary to specialization led the Committee on Tests and Measurement of the American Council on Education in 1932 to construct and offer to colleges a group of tests to be used in what they called "the sophomore testing program." Objective examinations of a comprehensive nature were constructed in an effort to measure general achievement and thus to provide a valuable supplement to course grades and hours of work in estimating the general level of students' academic accomplishment.

The sophomore testing program at the University of Minnesota was initiated in the spring of 1932. Since that time sophomores expecting to enter the Upper Division[1] have been tested regularly. The test results have been put to various uses. The Honors Committee has made some use of these records in deciding whether or not to admit a candidate to examination for higher honors. No systematic attempt, however, has been made to determine how well those test scores will differentiate honor students from those rejected by the committee. Nor has any attempt been made to determine the relative weights to be assigned to the tests in order to select students on the basis of test performance.

* This study was done at the University of Minnesota under the direction of Professor Donald G. Paterson. Parts II and III will appear in later issues of this journal.

[1] The term "Upper Division" refers to the second two years in the College of Science, Literature and the Arts.

The University Testing Bureau has found the results of this test battery valuable in advising many students. Here the tests have been used in the evaluation of the students' educational backgrounds, although no attempt has been made to use the tests in the statistical prediction of future educational accomplishment. The tests have also been used to discover relative strengths and weaknesses. It was assumed that such information would be useful in differential guidance of students.

During these years the sophomore testing program has been on an experimental basis, and a need has been felt for more information about the value of these tests. Although many members of the faculty have used the test results in advising students choosing a major sequence, it has not been determined just how well these tests will differentiate successful students in various fields. Further, no attempt has been made to determine the relationship between these test scores and success in the Upper Division.

The purpose of the present investigation has been to provide answers to the following three questions: First, will these tests differentiate successful major groups? Second, can a common minimum level of performance for all major groups be determined for admission into the Upper Division which will reduce the incidence of failure in these last two years? Third, how well will these test results differentiate students who receive degrees *magna cum laude* and *summa cum laude* from all those applying for degrees with higher honors?

This study is not an attempt to make a complete evaluation of the sophomore test battery. Nor is it possible to test all the hypotheses suggested by the three questions stated as the purpose of this research. It is hoped that the quantitative approach undertaken may indicate valuable ways of attacking such problems, and that it will yield results which may be of some practical significance for counselors and advisers faced with the difficult task of helping the student to select a major in harmony with his abilities and achievements. In this article

the material covered is confined to that related to the first of the three questions stated as the purpose of this research. Papers dealing with other aspects of the total problem will appear later.

The subjects included in this study were all the students who took the complete sophomore test battery in the years 1932–33, 1933–34, 1934–35, 1935–36. While almost all students took the entire group of tests in the sophomore program, some had incomplete records and were, therefore, not included in this study. Those taking the tests in the first year the battery was administered were not included in this study because the General Culture Test for that year was not strictly comparable to later forms of the test.

The complete sophomore test battery at the University of Minnesota included the Cooperative General Culture Test, the Cooperative English Test, and the Cooperative General Science Test. Those taking the battery during the first three years of those considered in this study were also given the Cooperative Literary Acquaintance Test.[2] The reliability coefficients for the 1935 form of these tests as reported by the Cooperative Test Service range from .93 to .98. Only slight variation is observed in those coefficients when corrected for the range of talent observed in major groups.

The test scores for students tested in the four-year period covered by this investigation were collected and equated in terms of the 1934 form of the tests in order to make all raw scores directly comparable. This was done by means of tables furnished by the Cooperative Test Service. In equating all three parts of the English test and the literary acquaintance test, "Scaled Score" values also furnished by the Cooperative Test Service were assigned to the raw scores in order to make different years comparable.

The University record of each student was obtained from the Registrar's office. The number of hours of A, B, C, D, and

[2] These tests are published by the Cooperative Test Service of the American Council on Education, New York City.

F were calculated for each student for his first two-year period and for his second two-year period, and honor point ratios were calculated for each period.[3]

In classifying students according to major groups it was next necessary to make certain arbitrary combinations. The number of cases for majors in some fields was not sufficient to warrant individual treatment. Majors in sociology, economics, political science, geography and anthropology were included in one group designated as social science. Majors in zoology, physics, chemistry, bacteriology, botany, anatomy and geology were combined into a natural science group. The group designated as liberal arts contains not only majors in liberal arts but also majors in philosophy and students from the University College. Majors in liberal arts and students from the University College present a rather wide range of educational interests. The liberal arts curriculum was created to enable students to attain the ends of higher education, allowing each one to do the work best suited to develop his powers and satisfy his interests. The University College arranges special courses for individual students whose intellectual and professional aims cannot be satisfied by other colleges of the University. Composition majors were combined with English majors, and all majors in the Romance languages were combined into one group. Mathematics was combined with astronomy. All the remaining major groups were treated individually, although some of these had relatively few cases. The number of cases was small for majors in pre-medicine, fine arts, elementary education, architecture and music, but large differences between these and other major groups were apparent. Other major groups treated individually were journalism, speech, library training, psychology, business administration, medical technology, law, history, diplomacy and consular service, and B.S. in social work. Thus, there was a total of twenty-one groups.

[3] Honor point ratio is an average grade calculated by assigning 3 points for one hour of A, 2 points for one hour of B, and one point for one hour of C, no points for a grade of D, and −1 for one hour of F, and dividing the sum by the number of hours.

Before it was possible to consider the three main questions of this investigation, it was necessary to discover whether or not performance for major groups has remained constant from year to year. If the performance of some major groups is not consistent from year to year, guidance on the basis of these tests or predictions based upon previous results cannot be made with as much confidence as if performance is consistent from year to year. If changes on some tests are found to exist, it is necessary to find which major groups show such changes. Since national norms were used in equating different forms of the tests any changes observed would refer to a particular major group as compared with sophomore performance in general. The comparison of performance for different years was limited to the total general culture test and to the total English test.

Since it was necessary to consider each major group separately, the number of comparisons would have been excessively large if each separate test in the sophomore battery had been considered. The answer to such questions is usually sought by comparing the mean performance of each year with every other year to observe any change that has taken place. In order to short-cut the large number of comparisons necessary, the same result has been achieved by a different method. An analysis of variance between years was calculated for each major group for both the total General Culture Test and the total English Test. By this method it was possible to test the hypothesis that the means for different years may be considered samples from the same population.

It was found that evidence of change in performance for major groups between different years never exceeded the upper level of significance (a probability value of .01). Students in the same major groups tested in different years were then combined since the observations were in accord with the hypothesis that they come from the same population. Tests of the homogeneity of variance revealed that major groups do not differ significantly with respect to the amount of their variability.

The application of various statistics herein employed carries the assumption that the distribution of scores is normal in form. Although this is frequently not a limiting assumption, it is advisable to observe the nature of the distributions before proceeding further. Rather wide departures from the normal are usually tolerated in the application of the statistics used in this study.

In order to discover the nature of the distributions of scores on these tests, eighteen distributions selected from all those available were plotted and tests for their deviation from the normal were made. The statistics g_1 and g_2 were calculated for measures of skewness and kurtosis. The results of these tests indicated that there is a tendency for the distributions to be positively skewed although most of these values are statistically non-significant. Three of the statistics yielded a probability value of less than .01. Less consistency was observed with respect to kurtosis although two distributions indicated significant leptokurtosis. In no case did a single test show significant skewness for all major groups. Tests which are closely related to a major field yield distributions which more closely approximate the normal for that major group than do tests which are not obviously related to the major field.

The value of these tests as instruments of guidance depends first of all upon the degree to which they will differentiate successful students in various fields of work. Before it is possible to advise a student to enter a particular major on the basis of test performance, it is essential to know how majors successful in that field have performed on the same tests. Estimates of relative strengths and weaknesses should be made not in terms of the performance of all sophomores but should be made relative to the performance of successful students in that major group. Until we know how well various major groups perform, the guidance value of these tests is severely limited.

We must first discover then whether or not different major groups can be distinguished on the basis of level of perform-

ance on the various sophomore tests. In order to answer this question, the performance of major groups was compared, with regard to both test performance and the second two years of honor point ratio by means of analysis of variance. The results of these analyses are given in Table 1.

TABLE 1

Results of Analyses of Variance between Major Groups

Measure	F Value	Probability
History and Social Science	7.07	$< .01$
Foreign Literature	7.03	$< .01$
Fine Arts	7.73	$< .01$
Total General Culture	3.85	$< .01$
English Usage	3.29	$< .01$
Spelling	3.00	$< .01$
Vocabulary	4.79	$< .01$
Total English	3.22	$< .01$
General Science	48.21	$< .01$
Second 2 yrs. Honor Point Ratio	6.59	$< .01$

These results indicate that with respect to no test can major groups be considered samples from the same population. Further, a difference between major groups is indicated for the second two years of honor point ratio. In order for these results to be valid as an indication of mean variation between major groups and to justify the use of the analysis of variance on these data, tests of the homogeneity of the variance were made for each test. These results indicated that the variances within major groups can be considered samples from the same population for all tests except general science. We, therefore, conclude that major groups do show significant variation in mean performance on all measures except the general science test since no part of the difference between groups can be charged to heterogeneity of variance.

The significant heterogeneity for the general science test indicates that at least some major groups differ significantly with respect to variability on this measure. In comparing

mean performances of any two or more major groups on this test, it is, therefore, necessary to consider the possibility that they also differ with respect to variability. Individual tests of significance, however, reveal highly significant differences between means of groups for groups with non-significant differences in variability.

The value of these tests for guidance depends upon the relationship between test performance and scholastic success in the Upper Division as well as upon the variability between major groups. It is necessary then to consider the relationship between these tests and scholastic success.

The correlation between the tests and the second two years of honor point ratio was calculated separately for each major group. These correlations are reported in Table 2. The r's

TABLE 2

Correlations between Test Scores and Second Two Years Honor Point Ratio for Certain Major Groups

Major group	N	History and Soc. Sci.	Foreign Lit.	Fine Arts	Eng. Usage	Spell.	Vocab.	Gen'l Science
English	82	.43*	.43*	.30*	.30*	.31*	.38*	.08
Journalism	89	.45*	.13	.14	.13	.18	.31*	.36*
Romance Language	44	.44*	.48*	.19	.34	.44*	.36	.28
History	29	.72*	.59*	.54*	.65*	.36	.57*	.03
Social Science	85	.42*	.30*	.14	.27*	.15	.32*	.29*
B.S. in Social Work	163	.40*	.30*	.24*	.25*	.28*	.20*	.22*
Natural Science	65	.19	.27	.29	.42*	.41*	.30	.28
Liberal Arts	83	.19	.32*	.30*	.35*	.15	.39*	.36*
Library Training	41	.41*	.31	.19	.23	.08	.05	.15
Music	17	.47	.70*	.63*	.75*	.36	.66*	.58
Business	23	.42	.08	.03	.35	.03	.37	.13
Psychology	24	.39	.38	.20	.54*	.12	.40	.33
Speech	24	.24	.57*	.36	.20	.15	.26	.19
Medical Technology	24	.47	.55	.39	.62*	.50	.36	.40
Diplo. & Cons. Serv.	17	.45	.03	.12	−.17	.11	.05	−.24
Mathematics	14	.48	.48	.27	.36	.26	.49	.70*

* Correlations significantly greater than zero.

which are significantly greater than zero have been starred. There is considerable variation in these correlations. For example, the correlations for the history and social science test range from .72 for majors in history to .19 for majors in natural science. Correlations for the English usage test range from .75 for majors in music to −.17 for majors in diplomacy and consular service. None of the tests yields a significant relation to success for all major groups and no test yields a relationship which is non-significant for all groups. The number of cases involved in some of the groups is rather small, and large variations in the relationship may be expected. For this reason two groups do not yield significant correlations for any of the tests although some of these correlations are high.

These correlations clearly indicate that the problem of the prediction of scholastic success should be approached by treating major groups separately. When major groups are combined, an averaging effect of these relationships is produced for a given test, and thus the value of the test is reduced for some major groups and a spurious estimate of its value is given for others. One would hardly expect that a test designed to measure knowledge in history and the social sciences would yield the same relationship to success for majors in the natural sciences as it does to success for majors in the social sciences or history.

We have already seen that major groups differ significantly in their mean performance on these tests. Furthermore, differences between major groups with respect to honor point ratio for the second two years was observed. These observations furnish additional indications of the necessity for the separate consideration of the prediction of scholastic success for various major groups.

Before it was possible to determine which major groups cannot be differentiated on the basis of test performances, individual comparisons were necessary. In order to facilitate the numerous comparisons necessary, a profile was drawn for each major group in terms of its deviations from the mean of

all majors who graduated. The limits within which these means could be expected to vary were then indicated on each profile. A probability level of .01 was used for these comparisons. By the comparison of these profiles, it was then possible to secure estimates of the significance of the difference between pairs of major groups. Such comparisons were used only in a rough way to indicate where individual tests of significance should be calculated. Differences in correlation coefficients were also considered in combining various major groups.

These profiles indicated a striking similarity of performance among majors in journalism, social science, liberal arts, law and business. An analysis of variance between these groups on each of the tests showed that the observations were in accord with the hypothesis that they could be considered samples from the same population. These major groups were therefore combined into a single group. Since majors in speech and library training also showed non-significant differences, they were also combined.

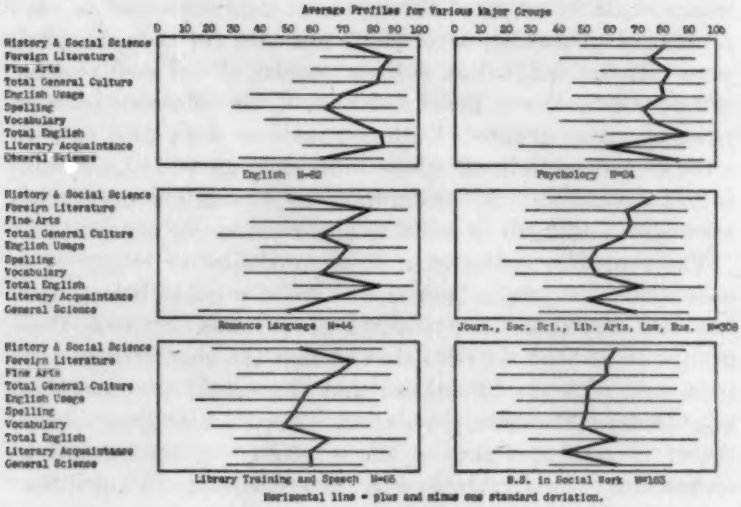
All the remaining major groups gave evidence of significant deviation among themselves and have, therefore, been considered as separate groups.

Profiles were then drawn for each group that could be differentiated on the basis of test performance. The means of the groups are expressed in terms of percentiles calculated from the performance of all sophomores who took the tests in the four-year period covered by this study. These are shown in Figures 1 and 2.

It is observed that the profiles for majors in English, Romance language and the combined group of library training and speech are similar in appearance but seem to represent different levels of performance on all the tests. Similar patterns are also observed for psychology majors, the combined group of majors in journalism, social science, liberal arts, law, and business, and majors in social work. These groups also seem to represent a hierarchy of similar performance. This tendency is less marked but nonetheless noticeable among

Figure I

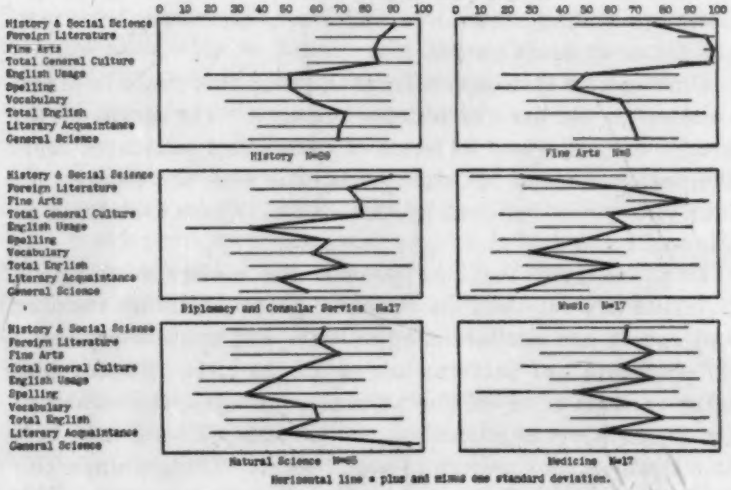
Average Profiles for Various Major Groups



Horizontal line = plus and minus one standard deviation.

Figure II

Average Profiles for Various Major Groups



Horizontal line = plus and minus one standard deviation.

majors in pre-medicine, natural science and medical technology.

Adequate discriminations may be difficult to make between major groups in these hierarchies, but differentiations between groups involving different patterns of ability should be far easier. The importance of making judgments of relative strengths and weaknesses in terms of the major group under consideration is easily understood when one realizes that a point one standard deviation below the mean of pre-medicine majors on the general science test is at the ninety-third percentile in terms of the performance of all sophomores. Few counselors would estimate that even when compared with majors in pre-medicine a student who obtained a score as high as the ninety-third percentile would be below approximately eighty-four per cent of successful students with whom he would have to compete if he followed the pre-medical curriculum. The lowest score obtained by any major in this group was at the eighty-third percentile.

From these profiles, it can be seen that major groups differ widely in their performance on these tests. Just how well students can be classified according to major group on the basis of test performance, however, is not apparent. An application of a method of approach to this problem will be discussed in the next article to appear in this series.

EXPERIMENTAL AND STATISTICAL ANALYSIS OF THE EFFECTIVENESS OF DEODORANT CREAMS

T. A. JACKSON, E. A. JEROME, AND N. SCHOENFELD
Columbia University

PROBLEMS in applied psychology are frequently baffling to the experimentalist because the methods at his command are not suitable. In many cases a new approach must be developed or an old one modified considerably. Something of this type of difficulty was met in the present attempt to determine the relative effectiveness of deodorant creams. Many tentative approaches were discarded before arriving at the method described in this article.

Some general considerations kept in mind in undertaking this problem are: (1) The testing conditions should not be made so artificial that the findings would be unrelated to the experience of everyday life. (2) The results must be quantitative in order that comparisons may be clear-cut. (3) The data should be based upon tests made on women since they are the greatest users of these products. (4) The testing should be done in rooms of summer temperature since the problem of controlling perspiration odors is greatest at that season.

The problem as a whole will be discussed under three headings: (1) How the samples of perspiration were collected. (2) How quantitative measures of the intensity of the odor samples were obtained. (3) How the data were treated statistically.

The most feasible method of producing perspiration for the purpose of this investigation was considered to be vigorous exercise. A stationary bicycle was used for this, care being

taken to adjust it to the same wheel tension at each period. The "workers" peddled at a constant speed for twenty minutes continuously. This caused each one to perspire freely. In order to get a sample of the perspiration to be tested for odor intensity, small odorless gauze pads were tied under each arm. No clothing could touch the pads. They were weighed and placed under the armpits before exercise was begun, then removed and weighed immediately upon cessation of exercise. The net weights of the perspiration samples varied from 500 to 2000 milligrams. A woman who could not produce 500 milligrams of perspiration was not used for these tests. Unpublished data indicated that the secretion of perspiration from the two arms is about the same (90 per cent of the time they differ not more than 10 per cent). Since this was the case it was possible to allow each worker to act as her own control. That is, if application of a cream is made to one armpit and not to the other, the difference between the intensity of the two different odor samples would be an index of the effectiveness of the cream.

The program of experimentation for a parteiular "worker" consisted in: (1) (Preliminary) Working twenty minutes for three days on the exercycle to get herself in condition for the experiment proper. (2) Working for another three days. On each of these days thirty minutes before the exercising was begun a deodorant cream was applied to one armpit. In this thirty-minute interval during which the cream was taking effect, the worker rested in the experimental room, the temperature of which was 85–90° F. and the humidity was about 40 per cent. Samples of perspiration were taken as previously indicated. Since they were to be judged they were kept in air-tight jars at body temperature until such judging was done. This interval was never more than three hours. (3) Two days after completing the above program the workers again came to the laboratory for similar tests with another deodorant cream. With half of the women cream A was used in the first three-day test, and cream B was used in the second. With the other half

the creams were used in the reverse order. The amount of cream used for each application was carefully standardized. For a given subject the creams were always applied to the same arm. All other conditions of the experiment were identical for the two creams tested.

Ten young women were used as workers. They were all healthy and were between the ages of seventeen and twenty-five. They were not scheduled for work during the menstrual period.

No use will be made of the weight data in this report. It was necessary to collect it in order to be sure the samples stayed up to the 500-milligram level. A smaller sample than that is not desirable for use in the odor detection method to be described below.

The first method considered for estimating the intensity of the odor sample was the one used by Zwaardemaker. This was discarded because of the difficulty of preparing the stimulus cylinders. The Elsberg blast-injection technique was next tried but it too was rejected because in a preliminary try-out of the method it was found that with perspiration odors, large volumes of air had to be injected under high pressure before the odor could be detected. The method we finally adopted was the "stream" injection technique which Elsberg employed to induce sensory adaptation. This method allows of injection of large volumes of air under low pressure, and seems especially useful for weak stimuli such as those involved in this work.

The stimuli to be judged were gauze pads saturated with perspiration; each was in a 500-cc. glass jar. With the stream technique, a very mild stream (2500 cc. per min.) of pure odorless air was caused to flow into the stimulus jar. This air current flows through the odor-laden gauze and out of the jar through the subject's nosepiece. The air going into the judge's nose was, of course, more or less odorous. If the sample of perspiration were large (1500 mg.) and if no cream had been applied, a judge could usually detect (perceive) the

odor in ½ to 3 seconds. If the odor were very weak, as when a cream had been applied, the reaction time was greater. The index of cream efficiency used in this study is the difference between the reaction time to the application sample, and the reaction time to the control sample of a given worker. Hereafter, this value will be referred to as a "delay score."

At a single session a judge was tested on two samples—the one from the control arm and the one from the application arm. Four reactions were required to each. The two samples were mixed up in random order for presentation to the judge. On the basis of these eight reaction times one delay score was computed. Successive reactions were separated by at least 30 seconds.

RESULTS

All the scores for one of the judges are shown in Table 1. The trend of the results can be seen pretty well by inspection of this table.

The "total delay" score for cream A was greater than for cream B for every exerciser. The over-all "total delay" score for cream A is 6001 as compared with 4050 for cream B; the A score is 48 per cent greater than B. A cumulative effect through the three days is not present in cream A and probably not present in B.

This superficial survey of the results gives an orientation for a more refined analysis of the results. To determine the reliability of these results, the method of "analysis of variance" was applied instead of the commoner critical ratio method.

Table 2 shows a summary of results obtained by this method. This is a three criteria type of analysis. By it an evaluation of the difference between creams, the difference between days, and the difference between exercisers may be obtained. Similar examples of three criteria of classification analysis may be found in Goulden (pp. 131–135).

Space will not permit a full exposition of the steps of com-

TABLE 1

Showing All the "Delay Scores" for One Judge. (Scores are in terms of one tenth seconds)

		Ex1*	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	A sums	B sums
1st day	A†	161	241	198	190	196	197	187	216	230	197	2013	
	B	139	143	112	152	146	118	114	85	146	41		1196
2nd day	A	168	206	137	201	121	198	217	140	262	255	1905	
	B	127	198	139	216	139	151	155	150	120	134		1529
3rd day	A	272	235	187	180	166	232	204	173	227	207	2083	
	B	94	117	196	127	121	148	122	131	137	132		1325
A sums		601	682	522	571	483	627	608	529	719	659	6001	
B sums		360	458	447	495	406	417	391	366	403	307		4050
A & B sums		961	1140	969	1066	889	1044	999	895	1122	966	10051	

* Ex1, Ex2, etc., refer to the different exercisers.

† "A" & "B" refer to the two deodorant creams used.

puting the F values shown in Table 2. However, to evaluate them (for creams, days or exercisers) they are compared with certain values obtained from standard tables for these particular conditions (degrees of freedom). If an F value were just equal to the 1 per cent value then it could be said that the obtained results could have arisen by chance only once in a hundred times.

The F value for "creams" in Table 2 is so high that such

TABLE 2

Showing the Summary Table of the Analysis

Source of variation	Sum of squares	DF	Variance* (Mean square)	F†	Criteria of significance	
					5%	1%
Creams	63440.0	1	63440.0	61.01	4.41	8.28
Days	1515.1	2	757.6	.73		
Exercisers	11250.2	9	1250.0	1.20	2.46	3.60
Creams × days	5731.4	2	2865.7	2.76	3.55	6.01
Exercisers × creams	14410.8	9	1601.2	1.54	2.46	3.60
Exercisers × days	17046.2	18	947.0	.91		
Triple interaction or remainder	18717.3	18	1039.9			
Totals	132111.0	59				

* Found by dividing Sum of Squares by DF (degrees of freedom).

† F values are found by dividing the variance (for creams, days, exercisers, etc.) by the "triple interaction" variance.

results could have arisen by chance only one time in many thousands. For practical purposes this means certainty, that is, cream A is superior to cream B. The differences among "exercisers" are not significant. We can see that the creams do not have significantly different effects on different "users" by testing the exerciser variance against the exerciser × creams interaction. Thus, $F = 1250.0/1601.2 = .78$, which is not significant. The F for days is also not significant, which indicates

among other possibilities, that there was not a clear-cut cumulative effect of the creams from day to day.

A similar analysis not shown here was made on the results for another judge. The difference between creams was not as significant as in the present case, but it also favored cream A. It reached approximately a 6 per cent criterion, meaning that the result could have arisen by chance only six times in a hundred. The total delay scores for this second judge were 1594 for cream A, and 1274 for cream B. The A score is 25 per cent greater than the B score.

SUMMARY

To determine the relative effectiveness of two deodorant creams (A and B), ten workers were exercised vigorously on a stationary bicycle, using a cream deodorant under one arm and nothing under the other. A sample of the perspiration from each arm was obtained by tying gauze pads to the armpit before exercise was begun. These pads were removed after the workout and placed in air-tight jars for measurement of the intensity of the odor. Each of the exercisers worked six days, three on cream A and three on cream B.

The index of the intensity of the odor was based on reaction time to these stimuli when they were presented to a judge by having a mild stream of air pass over the odorous pad to the judge's nosepiece. The reaction time to the stronger stimuli were shorter than those to weaker stimuli. The effect of the creams was to increase the reaction time. The observers made four reactions to the samples from each arm for all exercisers.

"Delay scores" were computed by taking the difference between the average reaction time for the control arm and that for the application arm. The scores were entered into a three-way table and the "analysis of variance" method was applied.

The results for one judge showed that the delay scores were greater for cream A than for cream B; the respective totals for all subjects over the three days—6001 and 4050. The

analysis of variance shows that the difference is highly significant. Variation between exercisers and between days was not found to be significant.

LITERATURE CITED

ELSBERG, C. A., AND LEVY, I. I. A new and simple method of quantitative olfactometry. *Bull. Neurol. Inst. New York*, 1935, 4, 5.

GOULDEN, C. H. *Methods of Statistical Analysis*. New York: John Wiley & Sons, 1939.

ZWAARDEMAKER, H. L'Odorat. 1925.

A QUALITATIVE ANALYSIS OF THE KENT ORAL EMERGENCY TEST AS A CLINICAL INSTRUMENT IN THE EXAMINATION OF NAVAL RECRUITS[1]

ROBERT J. LEWINSKI

Lieutenant (jg) H–V (S), United States Naval Reserve
U. S. Naval Training Station, Great Lakes, Illinois

THE appearance of the Kent Oral Emergency Test (5) in 1932 marked a radical departure from current mental test practice and theory. Clinical psychology had accepted the justifiable position of Binet and others to the effect that a valid test of general intelligence should sample many intellectual factors. In so doing, time and patience were necessary and the administration of a mental test was no expeditious matter. Although Doctor Kent did not offer her test as a tool with which absolute diagnoses of mental ability could be made, it was presented as an instrument which might prove to be convenient and practical in emergency situations.

There have appeared during the past few years several studies having reference to the Kent test. Elwood, Burchard, and Teagarden (3) published a general evaluation of the examination which was followed by Teagarden's (7) article introducing a method of converting the test scores into mental age years and months. Benton (2) compared the Kent test and the revised Stanford-Binet test when administered to school children, and Hogan (4) conducted a somewhat similar study using both children and adults as subjects. More recently, Springer (6) obtained correlations between the Stan-

[1] The opinions and assertions contained in this paper are those of the writer and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

ford-Binet and Kent tests applied to adolescent delinquents. Of the many findings reported in the above studies, one of the most significant is that a high positive correlation seems to exist between the ratings obtained on the Kent and Stanford-Binet scales.

The present state of national emergency has led to widespread military conscription and increased naval recruiting. This has magnified the need of a valid test of general intelligence which does not have the defects of group tests, can be administered in a reasonably brief period of time, and which can be used advantageously in the detection of the mental defective. Experience gained during the first World War demonstrated that the mentally deficient individual was a detriment to the armed forces, and that he adversely influenced efficiency and morale. One of the primary tasks of the military psychologist today is, therefore, to eliminate the mentally deficient before they are finally accepted in any branch of the service.

A procedure has been established at the Great Lakes Naval Training Station to detect the mentally unfit and unstable through individual examination. Each incoming recruit is given a summary psychiatric interview, and during this interview an attempt is made to determine the existence of such defects as psychotic and psychopathic states, psychoneurotic tendencies, and mental retardation. Those recruits suspected of mental retardation are referred for psychometry. In compliance with such referrals, the Kent Oral Emergency Test has been employed extensively by the writer. It should be noted that this examination has seen previous use in certain army induction centers (1) and that it has been recommended for this purpose (9). No final diagnoses of mental ability are made on the basis of this test alone. Those men receiving suspiciously low scores are retested with the more penetrating Bellevue-Wechsler Adult Intelligence Scale (10). Correlations between these examinations are promisingly high at this time and will be reported in a later paper when sufficient data are available.

PURPOSE AND RESULTS

Certain shortcomings of the Kent test prompted this study. In the first place, no qualitative or quantitative analysis of possible responses to the various test items was published in the original monograph. Although suggestions were offered as to what constituted an acceptable response and what credit should be given, there was no explanation as to what one might expect in terms of borderline and unacceptable responses. It is obvious that the shorter the test, the more increasingly important do scoring techniques become. The fact that question-response analysis is significant is evidenced by the space afforded it in the manuals of Terman and Merrill (8) and Wechsler (10).

Secondly, no analysis of individual items of the test was presented demonstrating their diagnostic significance. The examination was constructed so that there were three overlapping scales and five different groups of questions, the last four of which were arranged in order of difficulty. Although these scales were graded so as to be applicable to varying levels of intelligence, it was not shown that the questions composing each scale actually differentiated these levels.

The purpose of this paper is to attempt to meet these shortcomings by presenting an analysis of individual items of the Kent test and actual responses to them. It is believed that such an analysis may prove of value to other clinicians contemplating its use in similar fields.

The subjects used in this research were 300 naval recruits referred for psychometric examination to determine their fitness for military duty. They were recruited from the various states comprising the Ninth Naval District. The chronological age range was from 17 to 38.5 years, with a mean chronological age of 19.1 years. The mean school grade completed was the eighth. The range in IQ's extended from 56 to 101 with a mean and median of 80.8 and 81 respectively.[2] The standard

[2] With the scoring techniques used, the highest obtainable mental age was 15 years and 1 month and the maximum IQ 101.

deviation of the mean was 9.44. Obtained intelligence quotients were distributed as follows: Normal (90+) 17.7%; Dull Normal (80–89) 37.3%; Borderline (70–79) 33.3%; Mentally Deficient (below 70) 11.7%. In computing mental ages from raw scores, the method suggested by Teagarden (7) was employed. In deriving intelligence quotients, a basal chronological age of fifteen was used. The recruits referred for psychometric examination were those suspected of psychologic disability and therefore the mean IQ reported does not reflect the intellectual level of the typical naval recruit. In the majority of cases (75%), the Middle Scale was used, although the Upper Scale was employed when warranted. It was unnecessary to use the Lower Scale, which is too elementary for the subjects tested. Questions 6–10 (Group II), therefore, will not be considered in this study.

Responses to the various questions were recorded verbatim. It was difficult to make clear-cut distinctions in many instances in terms of plus and minus credit. Some responses were considered as "borderline," and these were scored plus or minus depending on subsequent clarification afforded by the subject. The lists to follow contain actual responses of the subjects used in this study and are separated into plus, borderline, and minus answers as determined by the writer. In making this determination, the original suggestions of Kent were followed. These lists, while necessarily not all-inclusive, contain the responses that one might expect in examining a similar group of subjects. The responses are listed according to their frequency of appearance.

GROUP I

Question 1. *What are houses made of? Any materials you can think of.* (Value: One point for each item, up to four)

Plus responses: Any acceptable building material.

Borderline responses: (To be scored plus or minus depending on qualification given by subject) Marble, paper, rocks, blocks, asphalt, tar.

Minus responses: Hammers, saws, drills, water, chimneys, roofs, windows, floors, furnaces.

Question 2. *What is money made of?* (Value: One point for each item, up to five)

Plus responses: Silver, copper, paper, gold, nickel.
Borderline responses: (To be scored plus or minus depending on qualification given by subject) Metal, alloy.
Minus responses: Brass, bronze, lead, zinc, aluminum, platinum, tin, thread, rags, cloth, ink, glass.

Question 3. *Why does cork float on water?* (Value: One to three points, depending on quality of response)

Plus responses: (Three points) Density of water is greater. Water displaced is heavier than cork. Cork is less heavy than water displaced.
Plus responses: (Two points) Lighter than water. Water is heavier. Water outweighs cork.
Plus responses: (One point) Light. Lighter. "It's not as heavy."
Borderline responses: (To be scored plus or minus depending on qualification given by subject) "A difference in weight." "More buoyancy."
Minus responses: "It's made of wood." "A rubber substance." "Air can't get in." "Holes in it." "Won't soak up water." "It's made of bark."

Question 4. *What is sand used for?* (Value: One, two, or four points, depending on quality of response)

Plus responses: (Four points) Glass. "For making glass." "To manufacture glass."
Plus responses: (Two points) Cement, concrete, plaster, bricks, mortar.
Plus responses: (One point) High-jumping pits, sand boxes, sandpaper.
Minus responses: "To grow vegetables." "For farming."

Question 5. *If the flag floats to the south, from what direction is the wind?* (Value: Three points)

Plus response: North.
Minus response: Any deviating response.

GROUP III

Question 11. *How many minutes in an hour?* (Value: One point)

Plus response: Sixty.
Minus response: Any deviating response.

Question 12. *How many days in a month?* (Value: One to three points)

Plus responses: Thirty, thirty-one.
Borderline responses: (To be scored plus or minus depending on qualification given by subject) Twenty-eight, twenty-nine. (One additional point if subject refers specifically to February)
Minus responses: "Seven days in a week, twenty-eight in a month." "About twenty-eight."

Question 13. *Why is it colder at night than in the daytime?* (Value: One to three points, depending on quality of response)

Plus responses: (Three points) "The earth rotates and the sun disappears." "Sun's rays are absent because of the motion of the earth." "The earth's movement hides the sun's rays."
Plus responses: (Two points) Any response referring to the absence of the sun and not specifically mentioning the rotation of the earth.
Plus responses: (One point) This credit is given only when it is necessary to reverse the question so as to read: "What makes it warmer in the daytime than at night?"
Minus responses: "The winds are stronger." "It's darker at night." "There's less wind in the daytime." "It rains more at night."

Question 14. *Tell me the names of some birds.* (Value: One point each, up to four)

Plus responses: Any recognized bird.
Minus responses: Bat, butterfly.

Question 15. *Now the names of some fishes.* (Value: One point each, up to four)

Plus responses: Any recognized fish.
Borderline responses: (To be scored plus or minus depending on qualification given by subject) Minnow, chub.
Minus responses: Whale, oyster, lobster, shrimp.

GROUP IV

Question 16. *At what time of day is your shadow shortest?* (Value: Three points)

Plus responses: Noon, midday, twelve o'clock.
Borderline responses: (To be scored plus or minus depending on qualification given by subject) "When the sun's right above you." "When the sun is directly overhead."
Minus responses: "In the evening." "At supper time." "At sundown." "When you can't see it."

Question 17. *Give the names of some large cities.* (Value: One point each, up to four)

Plus responses: Any city with a population in excess of 100,000.
Minus responses: New York, California, Pennsylvania, Texas, New Jersey, Florida, Grand Forks, Belmont, Ohio, Phoenix, Joliet, Colorado, Oregon, Canada, Old Mexico.

Question 18. *What is paper made of?* (Value: One point for each item, up to three)

Plus responses: Wood, rags, pulp, cloth, wood-pulp, corn-stalks, straw.
Borderline responses: (To be scored plus or minus depending on quality of response) Fiber, fibrous material.
Minus responses: Trash, hair, paste.

Question 19. *How many stripes in the flag?* (Value: Two points)

Plus responses: Thirteen. "Seven red and six white."
Minus responses: Any deviating response.

Question 20. *Whose picture is on a two-cent stamp?* (Value: Two or three points, depending on quality of response)

Plus responses: (Two points) Washington. (One additional point for mention of special or commemorative issue)
Minus responses: Jefferson, Lincoln, Roosevelt, "The President." "Uncle Sam."

GROUP V

Question 21. *What is the difference between a cable and a chain?* (Value: One to four points, depending on quality of response)

Plus responses: (Four points) "A chain is made of links and a cable is made of many wires twisted together tight." "A cable is made of a lot of small wires woven together while a chain is made of links."
Plus responses: (Three points) "Cables are ropes made of wire and chains are made of links."
Plus responses: (Two points) "A chain has links and a cable is all one piece."
Plus responses: (One point) "A chain has links." "A cable don't have links."
Minus responses: "Chains are longer." "Chains are iron and cables are copper." "Chains are used for towing." "A cable is stronger than a chain."

Question 22. *Why does the moon look larger than the stars?* (Value: Two to four points, depending on quality of response)
Plus responses: (Four points) "Things look larger when they're closer to you, even if they're not." "A smaller thing will look larger than a bigger thing if it's closer to you."
Plus responses: (Three points) "It's nearer." "It's nearer to the earth." "It's closer."
Plus responses: (Two points) "It's lower." "It's lower down."
Minus responses: "The moon is larger." "Stars are planets." "The moon gives more light."

Question 23. *What metal is attracted by a magnet?* (Value: Two or four points, depending on quality of response)
Plus response: (Four points) Iron.
Plus response: (Two points) Steel.
Minus response: Any deviating response.

Question 24. *What is electricity used for?* (Value: One to four points, depending on quality of response)
Plus responses: (Four points) "To generate power for many uses." "To furnish power for things run by electricity."
Plus responses: (One point each) Light, heat, power.
Plus responses: (One additional point) Radios, streetcars, motors, machinery, generators, electric fans, pumps.
Minus responses: "For batteries." "It kills people if they get close to it." "To run through wires."

Question 25. *If your shadow points to the northeast, where is the sun?* (Value: Four points.)
Plus response: Southwest.
Minus responses: Any deviating response.

The only arbitrary standard set by the writer was in regard to Question 17, *Give the names of some large cities.* Credit was given when the city named had a population of 100,000 or more, and no credit was given when the population fell below this number. Other clinicians may find it desirable to alter this standard depending on the particular location where the test is being used. Following Kent's suggestion, the response "New York" was taken to mean the state unless specifically referred to as "New York City." This proved to be the most frequent minus response recorded, while "New York City," as a plus response, ranked in twelfth position.

Logically, if a question is to be of any diagnostic value, it should differentiate varying degrees of intelligence. To evaluate the internal consistency of the Kent test, a statistical analysis was made to show the frequency with which the various responses to each question were given by the normal, dull normal, borderline, and mentally deficient groups. These data are found in Tables 1 to 4. They were constructed to show the percentage of subjects within each group who scored from the maximum to zero points credit. The normal group is not considered in Table 2 because of the small number of subjects tested with the Middle Scale. In Table 4 only the normal and dull normal groups are represented since few borderline and no mentally deficient subjects were examined with the Higher Scale. Examination of these tables will give some indication as to the degree the questions differentiate among the various intellectual levels. In considering the evaluation of the individual questions to follow, it should be noted that the writer's subjective impressions, formed through frequent administration of the test, were used to reinforce the statistical findings.

Of the questions contained in Group I, Question 2, *What is*

money made of? and Question 3, *Why does cork float on water?* proved definitely superior in differentiating among all intellectual levels. Question 1, *What are houses made of?* discrimi-

TABLE 1

Frequency of Credits Earned, Group I. Questions 1–5
(In Per Cent)

Question	Points credit					
	5	4	3	2	1	0
1. (Value: 1–4)						
Normal	—	87	13	0	0	0
Dull Normal	—	67	30	1	2	0
Borderline	—	60	37	3	0	0
Mentally Deficient	—	34	34	32	0	0
2. (Value: 1–5)						
Normal	26	62	12	0	0	0
Dull Normal	3	51	38	5	2	1
Borderline	2	35	40	14	7	2
Mentally Deficient	0	12	49	9	26	4
3. (Value: 1–3)						
Normal	—	—	9	64	25	2
Dull Normal	—	—	2	37	49	12
Borderline	—	—	0	11	69	20
Mentally Deficient	—	—	0	2	49	49
4. (Value: 1, 2, 4)						
Normal	—	45	—	55	0	0
Dull Normal	—	22	—	74	4	0
Borderline	—	9	—	84	6	1
Mentally Deficient	—	9	—	85	2	4
5. (Value: 3)						
Normal	—	—	100	—	—	0
Dull Normal	—	—	98	—	—	2
Borderline	—	—	90	—	—	10
Mentally Deficient	—	—	43	—	—	57

nated between the normal and dull normal groups and between the borderline and mentally deficient groups, but made little differentiation between the dull normal and borderline groups. Question 4, *What is sand used for?* was satisfactory except at

the two lower levels. Question 5, *If the flag floats to the south, from what direction is the wind?* was shown to be too elementary and was of value only in differentiating the mentally deficient from the borderline group.

TABLE 2

Frequency of Credits Earned, Group III. Questions 11–15
(In Per Cent)

Question	Points Credit				
	4	3	2	1	0
11. (Value: 1)					
Dull Normal	.....	.....	.....	100	0
Borderline	.....	.....	.....	98	2
Mentally Deficient	.....	.....	.....	86	14
12. (Value: 1–3)					
Dull Normal	.....	45	42	12	1
Borderline	.....	14	53	24	9
Mentally Deficient	.....	9	49	26	16
13. (Value: 1–3)					
Dull Normal	.....	2	94	2	2
Borderline	.....	0	85	4	11
Mentally Deficient	.....	0	63	9	28
14. (Value: 1–4)					
Dull Normal	100	0	0	0	0
Borderline	100	0	0	0	0
Mentally Deficient	94	4	2	0	0
15. (Value: 1–4)					
Dull Normal	96	4	0	0	0
Borderline	79	12	9	0	0
Mentally Deficient	46	26	20	8	0

Group III contained two questions of doubtful value. Question 11, *How many minutes in an hour?* was much too easy for the dull normal and borderline groups, as it was answered correctly by 100% and 98% of the subjects within these groups respectively. The weakness of this item is perhaps due in part to the fact that its answer is a matter of formal learning in

many school systems. Question 14, *Tell me the names of some birds,* afforded no differentiation between the dull normal and

TABLE 3

Frequency of Credits Earned, Group IV. Questions 16–20
(In Per Cent)

Question	Points Credit				
	4	3	2	1	0
16. (Value: 3)					
Normal	—	98	—	—	2
Dull Normal	—	93	—	—	7
Borderline	—	63	—	—	37
Mentally Deficient	—	18	—	—	82
17. (Value: 1–4)					
Normal	98	2	0	0	0
Dull Normal	80	17	3	0	0
Borderline	50	27	15	6	2
Mentally Deficient	26	24	29	15	6
18. (Value: 1–3)					
Normal	—	6	57	37	0
Dull Normal	—	2	47	49	2
Borderline	—	0	33	63	4
Mentally Deficient	—	0	6	68	26
19. (Value: 2)					
Normal	—	—	94	—	6
Dull Normal	—	—	84	—	16
Borderline	—	—	52	—	48
Mentally Deficient	—	—	32	—	68
20. (Value: 2, 3)					
Normal	—	0	44	—	56
Dull Normal	—	0	53	—	47
Borderline	—	0	41	—	59
Mentally Deficient	—	0	41	—	59

borderline groups and practically none between the borderline and mentally deficient groups. Satisfactory items in Group III were Question 12, *How many days in a month?* and Question 15, *Now the names of some fishes,* while Question 13, *Why*

is it colder at night than in the daytime? showed a weakness in discriminating at the dull normal and borderline levels.

In Group IV were found two highly satisfactory items, Question 17, *Give the names of some large cities,* and Question 18, *What is paper made of?* Question 19, *How many stripes in the flag?* discriminated well between the dull normal and borderline groups and between the borderline and mentally deficient groups. This question, while generally satisfactory in the examination of naval recruits, loses its value in testing men who have been in the service for any length of time, for here again the answer becomes a matter of formal learning. Question 16, *At what time of day is your shadow shortest?* was too easy for the two upper levels, but was excellent in discriminating the borderline from the mentally deficient group.

TABLE 4

Frequency of Credits Earned, Group V. Questions 21–25
(In Per Cent)

Question	Points credit				
	4	3	2	1	0
21. (Value: 1–4)					
Normal	8	46	42	4	0
Dull Normal	6	27	50	6	11
22. (Value: 2–4)					
Normal	0	66	24	—	10
Dull Normal	0	61	17	—	22
23. (Value: 2, 4)					
Normal	66	—	32	—	2
Dull Normal	50	—	33	—	17
24. (Value: 1–4)					
Normal	12	48	38	2	0
Dull Normal	6	33	61	0	0
25. (Value: 4)					
Normal	90	—	—	—	10
Dull Normal	61	—	—	—	39

Group IV contained the weakest question in the entire scale from a diagnostic standpoint, namely, Question 20, *Whose picture is on a two-cent stamp?* Only 44% of the normal group were able to answer this question correctly, while 53% of the dull normal group were able to do so. Forty-one per cent of the borderline and mentally deficient groups received credit on this item. Although it is possible to score a maximum of three points on this question, none of the subjects tested did so. There are two possible explanations of the weakness of this item. Firstly, postal regulations now require the use of a three-cent stamp for other than local mail whereas previously a two-cent stamp would suffice. Since the question was obviously designed to measure acuity of observation, its value was automatically reduced when the object to be observed was limited in its frequency of appearance. Secondly, the numerous commemorative stamps that have appeared during the past ten years especially have tended to confuse the issue further. These data certainly suggest that this question should be seriously considered for elimination or revision. A question such as, *Whose picture is on a penny?* which presumably measures the same function, might prove to be a good substitute.

With regard to Group V, only the two upper levels need be considered. Question 25, *If your shadow points to the northeast, where is the sun?* and Question 21, *What is the difference between a cable and a chain?* proved to be the best in differentiating the normal from the dull normal group, while Question 24, *What is electricity used for?* and Question 23, *What metal is attracted by a magnet?* ranked next in order. Question 22, *Why does the moon look larger than the stars?* was of little value in discriminating at these levels.

DISCUSSION

The writer has attempted to meet certain shortcomings of the Kent Oral Emergency Test by presenting a qualitative analysis of individual test items. Responses to the various

questions were considered in terms of plus, borderline, and minus credit with a view toward providing an adequate scoring guide for others contemplating the use of the test in similar fields. The merit of the individual items was further considered statistically in terms of their value in discriminating at the normal, dull normal, borderline, and mentally deficient intellectual levels. It was shown that while many of the questions were satisfactory in this regard, several were without value. The findings warrant the recommendation that certain of the test items should be considered for revision or elimination.

As a clinical instrument in the examination of naval recruits, the Kent test has proved its worth. On the whole, it may be used to advantage in detecting the mental defective. It can be administered in a short period of time, which is an asset when large numbers of subjects must be examined in the most efficient manner. The fact that it is an oral examination eliminates many of the defects of group tests.

Subjective evaluation has indicated that the test was inoffensive to the majority of recruits used in this study. Many of the questions have a direct positive appeal and are met as a challenge. The attitude with which the test is met by the testee is dependent on several factors, the most important of which are the approach employed by the examiner and the circumstances under which the test is administered. It has been noted by the writer, for example, that those men who are anxious to get into the service respond more readily than those who are indifferent or who want a discharge. As a rule, subjects within the lower age brackets (17–20) display more interest in the examination than do the more mature recruits. Also, subjects of the borderline and dull normal groups evidence more enthusiasm than those testing at the normal level.

Without doubt, there are many induction centers and recruiting stations faced with the problem of eliminating the mentally deficient in the most efficient manner possible. To meet such a situation, the Kent test is recommended. Certain

precautionary measures should, however, be observed: (a) The method of administration should be flexible, and dependent upon the circumstances under which the test is administered; (b) the examination should be suspended if the subject becomes apprehensive or emotionally upset; (c) subjects should not be told their ratings in terms of either mental age or intelligence quotient; and (d) the Kent test should not be substituted for more extensive psychometric examinations if a precise measure of mental ability is desired.

REFERENCES

1. Atwell, C. R., Bloomberg, W., and Wells, F. L. Psychometrics at an army induction center. *New England Journal of Medicine*, 1941, 224, 898–899.
2. Benton, A. L. Performances of school children on the revised Stanford-Binet and the Kent E-G-Y Test. *Journal of Genetic Psychology*, 1938, 52, 395–400.
3. Elwood, M. I., Burchard, E. M. L., and Teagarden, F. M. An evaluation of the Kent Oral Emergency Test. *Journal of Applied Psychology*, 1937, 21, 75–84.
4. Hogan, H. P. Comparison of Stanford-Binet and Kent Oral Emergency scores. *Journal of Genetic Psychology*, 1941, 58, 151–159.
5. Kent, G. H. Oral test for emergency use in clinics. *Mental Measurement Monographs*, January, 1932, No. 9, 50 pp.
6. Springer, N. N. Kent Oral Emergency and Stanford-Binet tests applied to adolescent delinquents. *American Journal of Orthopsychiatry*, 1941, 11, 292–299.
7. Teagarden, F. M. A method of interpolating Kent Oral Emergency Test scores into mental age years and months. *Journal of Applied Psychology*, 1937, 21, 468–473.
8. Terman, L. M., and Merrill, M. A. *Measuring Intelligence.* New York: Houghton Mifflin, 1937.
9. von Storch, T. J. C., and Pratt, G. O. Neuropsychiatric examination for induction or enlistment in the army. *The Military Surgeon*, 1941, 88, 627–633.
10. Wechsler, D. *The Measurement of Adult Intelligence.* Baltimore: Williams and Wilkins, 1941.

ESTIMATING THE MINIMAL RELIABILITY OF A TOTAL TEST FROM THE INTER-CORRELATIONS AMONG, AND THE STANDARD DEVIATIONS OF, THE COMPONENT PARTS

EDWIN E. GHISELLI
University of California

WHEN conditions are such that only one administration of a test is possible and a measure of its reliability is desired, the common procedure is to estimate it by means of the Spearman-Brown correction of the coefficient of correlation between the scores on two equivalent halves of the test. Occasionally, however, due to the nature of a test it is impossible to divide it into two equivalent parts. This situation is particularly true with many of the tests of manual dexterity and motor ability where the score on the total test is the total time, or the total number of items completed on the several constituent parts of the test. The different parts of these tests will not be comparable due to practice or fatigue effects, or due to the varying nature of the tasks required by the parts. Consequently, the Spearman-Brown correction formula cannot be employed since the basic assumptions underlying it, namely, that the parts be comparable in nature and that their standard deviations, and the intercorrelations among them and their reliability coefficients be equal, are not satisfied.

Recently the writer was confronted with this type of situation wherein a knowledge of the reliability of the test concerned was of considerable importance. Because of the expense involved in the rental of the O'Connor finger dexterity test a

small counseling agency desired a substitute, and proposed using a simple pegboard test of the type which can be made up at very small cost. However, it was necessary to discover the magnitude of the correlation between the scores on the pegboard and O'Connor tests in order to determine the extent to which they were measuring the same abilities. The pegboard test that was suggested was similar to the one described by Hayes,[1] since he reported its effectiveness in the selection of bench workers assembling small electrical parts to be about as great as that of the O'Connor test, the validity coefficients of both tests being of the order of .30 to .40.

To discover the degree of relationship between the scores on the two tests, a pegboard test was constructed following Hayes' design, and it and the O'Connor test were administered to a group of subjects. The pegboard test consisted of two rows of 25 holes each, and following Hayes' procedure the test was divided into three parts, insertion of pegs one at a time with the right hand, then with the left hand, and finally with both hands simultaneously. The score for the total test was taken as the sum of the scores on the three separate parts.

The coefficient of correlation between the scores on the pegboard test and the scores on the O'Connor test was found to be only .26. This meant, then, that the former test was inadequate as a direct substitute for the latter. The writer suggested that the low correlation between the two tests might be due to the fact that the reliability of the pegboard was low, and that perhaps by lengthening the test it could be made more reliable and its correlation with the O'Connor test consequently increased. Published data have indicated that the O'Connor test has a substantially high reliability, and therefore a lack of self-consistency on its part probably was not the determining factor in the low inter-test correlation. Since the reliability of the pegboard test at its present length could not be estimated

[1] E. G. Hayes, "Selecting women for shop work," *Person. J.*, 1932, 11, 69–85.

by the Spearman-Brown formula in the usual fashion due to a lack of equivalence among the parts, it meant, that unless some other means could be devised to estimate its reliability, another group of subjects would have to be run with the test in lengthened form. To solve this problem the writer devised the following method for estimating the minimal reliability of tests similar to the pegboard test employed here, which consist of parts that are not directly comparable.

It has been shown that the reliability coefficient of a total test can be calculated from the reliability coefficients of its component parts, their standard deviations, and the intercorrelations among them.[2] The general formula is

$$r_{tt} = \frac{\sigma_1^2 r_{11} + \cdots + \sigma_n^2 r_{nn} + 2\sigma_1\sigma_2 r_{12} + \cdots + 2\sigma_1\sigma_n r_{1n} + \cdots + 2\sigma_{(n-1)}\sigma_n r_{(n-1)n}}{\sigma_1^2 + \cdots + \sigma_n^2 + 2\sigma_1\sigma_2 r_{12} + \cdots + 2\sigma_1\sigma_n r_{1n} + \cdots + 2\sigma_{(n-1)}\sigma_n r_{(n-1)n}} \quad (1)$$

where r_{tt} is the reliability coefficient of the total test, r_{11} to r_{nn} the reliability coefficients of the n component parts, r_{12} to $r_{(n-1)n}$ the intercorrelations among the various parts, and σ_1 to σ_n the standard deviations of the parts. If the standard deviations of all of the parts are equal, and the intercorrelations among and the reliability coefficients of them are also equal, the formula reduces to

$$r_{tt} = \frac{n\sigma^2 r + n(n-1)\sigma^2 r}{n\sigma^2 + n(n-1)\sigma^2 r} = \frac{nr}{1 + (n-1)r} \quad (2)$$

which is, of course, the Spearman-Brown correction formula. Thus the Spearman-Brown formula is a special case of a more general formula for estimating the reliability of a total test from certain known characteristics of its component parts.

Formula (1) is the one which would be used to estimate the reliability of a total test when the component parts have unequal variabilities, and when the intercorrelations among them and their reliability coefficients are of different magnitudes.

[2] E. E. Ghiselli and G. Kuznets, "Short-cut methods for calculating raw and corrected correlations between a composite variable and its components," *J. Educ. Psychol.*, 1937, 28, 237–240.

However, in order to apply formula (1) it is necessary to know the reliability of each of the component parts. Since this information is not available from a single repetition of a test, it is sometimes necessary to estimate these reliability coefficients in order to substitute appropriate values in the formula.

It is impossible to estimate the exact magnitude of the reliability of a test if the only information is the extent to which it is related to one or more other tests, but it is possible to estimate the lowest value it can possibly have. If these minimum reliability coefficients are available and are used in formula (1) then the minimum reliability coefficient of the total test can be calculated. The highest possible correlation which can exist between two variables 1 and 2, is given in the formula

$$r_{1_\infty 2_\infty} = \sqrt{r_{11} r_{22}}$$

Knowing the value of this highest possible correlation and the reliability coefficient of variable 2, the reliability coefficient of variable 1 can be calculated, as

$$r_{11} = \frac{r^2_{1_\infty 2_\infty}}{r_{22}} \qquad (3)$$

In our case, however, we know neither the value of this highest possible correlation nor the reliability coefficient of variable 2. It is clear from formula (3) that with a given value of $r_{1_\infty 2_\infty}$, the higher the value of r_{22} the lower will be the value r_{11}. If we know $r_{1_\infty 2_\infty}$, and not r_{22}, but set the value of the latter equal to 1.00, a value would be obtained below which r_{11} could not possibly fall. The value obtained, therefore, would be the lowest possible value that the reliability coefficient of variable 1 could possibly have. Formula (3), then, resolves to

$$r_{11} = r^2_{1_\infty 2_\infty} \qquad (4)$$

It remains to estimate $r_{1_\infty 2_\infty}$. Since $r_{1_\infty 2_\infty}$ cannot be lower than r_{12}, this last coefficient will be our best estimate of the lowest possible value for $r_{1_\infty 2_\infty}$. Formula (4), then resolves to

$$r_{11} = r^2_{12} \qquad (5)$$

Our resulting estimate of r_{11} in most cases will be lower than

its true value, but we can be assured that the true value will not be lower than our estimate. The most effective value of r_{12} to use will be the coefficient of correlation of highest value between variable 1 and whatever other variable is under consideration. The square of this value can be used in formula (1), and the resulting value of r_{tt} which is obtained will be the minimal reliability coefficient of the total test.

We may illustrate the use of this method with data obtained from the pegboard test described earlier. The standard deviations of the three parts of the test were found to be $\sigma_1 = 5.1$, $\sigma_2 = 4.0$, and $\sigma_3 = 7.3$; the intercorrelations among the three component parts being $r_{12} = .75$, $r_{13} = .78$, and $r_{23} = .92$. The highest coefficient of correlation which exists between part 1 and either of the other two parts is .78. Consequently, the square of .78 is used as the estimate of r_{11}. The highest correlation which exists between part 2 and either of the other two parts is .92. The square of .92, therefore, is used as the estimate of r_{22}. Since .92 is the highest correlation between part 3 and either of the other two parts, the square of .92 is also used as the reliability of part 3. When these estimated reliability coefficients, together with the standard deviations of the three parts and the intercorrelations among them, are used in formula (1), the reliability of the total test is found not to be lower than .91.

Knowing that the reliability coefficient of the total pegboard test was at least .91, we were assured that its self consistency was adequate. It followed, therefore, that the relatively low correlation between the scores on the pegboard and O'Connor finger dexterity tests could not be explained on the basis of low reliability on the part of the former. Consequently, increasing the length of the pegboard test would not result in an appreciable increase in the value of this coefficient of correlation. Thus the above statistical procedure saved the labor of trying out the pegboard test in lengthened form with another group of subjects.

When the value of a statistical constant is desired but it is possible only to estimate it, the estimate will increase in usefulness as it approaches more closely the true value. It is desirable, therefore, to know the conditions under which an estimate will vary markedly from the true value. The estimated reliability coefficient of a total test computed by the method described above will become lower with a reduction in size of the intercorrelations among the component parts, with a decrease in the number of component parts, and with an increase in the size of the standard deviations of those parts which show the lowest correlations with other parts. When these conditions are adverse, the estimate of the reliability coefficient of the total test will be rather low, and since all that the method gives is the lowest possible reliability coefficient, it will have little value under these circumstances.

A STUDY OF THE INTERRELATIONSHIPS BETWEEN THE ADAMS–LEPLEY PERSONAL AUDIT AND THE BERNREUTER PERSONALITY INVENTORY

WILLIAM R. TUBBS*

THE past decade has witnessed an increasing interest in the development of instruments designed to measure personality characteristics or "traits." The development of factor analysis has opened new avenues of approach to the statistical analysis of personality. Business and industry, in seeking more objective measurement of personality, have contributed their share to the stimulation of interest in the field.

Dr. C. R. Adams and Dr. W. M. Lepley, of the Pennsylvania State College, have recently devised a form intended primarily for use in selection in industry. It is only natural that this form will be compared extensively with others in the field. This paper is a report of one such study.

THE BERNREUTER PERSONALITY INVENTORY

The Bernreuter Personality Inventory (2) provided psychologists with a radical departure from the then conventional tests for each "trait," since it purported to measure in one blank of 125 questions, four aspects of personality. Subsequent study by Flanagan (6), using factor analysis, showed that the Inventory measured two relatively independent traits. Scoring scales have been prepared which measure the traits, designated as follows: B1–N, Neurotic tendency; B2–S, Self-

* An abstract of a thesis submitted in partial fulfillment of the requirements for the degree, Master of Science, Pennsylvania State College.

sufficiency; B3–I, Introversion-extroversion; B4–D, Dominance; F1–C, Self-confidence; and F2–S, Sociability. In this study, B3–I was omitted from consideration because it has been found to be almost identical to B1–N (3).

Studies in the validity and reliability of the Bernreuter Personality Inventory are many. Bernreuter reports his own validities and reliabilities in the Manual.

THE ADAMS-LEPLEY PERSONAL AUDIT (FORM D)

This instrument is in the final stages of construction. The present form is the fourth revision. The first two forms were largely experimental and exploratory. The third form consisted of ten sub-tests. All but one of these is included in the fourth form. Reliability coefficients of the ten parts of the third form, obtained by the retest method are shown in Table 1.

TABLE 1

Reliability Coefficients on Form C of the Adams-Lepley Personal Audit Obtained by the Retest Method

Part	N	Reliability coefficient
I	32	.97
II	32	.96
III	32	.99
IV	32	.96
V	32	.94
VI	32	.95
VII	32	.94
A	42	.72
B	47	.72
C	42	.78

The cases in the above treatment were supplemented by other cases to compound a total of 356. These individuals were college students, male and female. These 356 cases were then used for a part by part item analysis, which was executed according to the technique of Sletto (8). On the basis of this

item analysis certain items were discarded from each part. The parts were then increased in length by adding new items until each sub-test contained sixty items. This achieved the present form (D) of the instrument. This form has been subjected to a second item analysis similar to the above mentioned. This last analysis, based upon the quartiles of a population of 400 cases (college students, both male and female) provided the basis for scoring used in the present study. The intercorrelations among the various sub-tests of this fourth form of the instrument are shown in Table 2.

TABLE 2

Intercorrelations of the Parts of Form D of the Adams-Lepley Personal Audit
N = 400

	II	III	IV	V	VI	VII	VIII	IX
I	−.06	.14	.40	−.07	.01	.01	.15	.02
II		.03	.09	.20	.11	.21	.01	.34
III			.17	.39	.04	.50	.08	.33
IV				.27	.18	.32	.17	.11
V					.19	.40	.03	.18
VI						.10	.08	−.01
VII							.13	.22
VIII								.04

One of the limitations in the use of personality measures in industry has been the difficulty experienced in getting applicants to reply truthfully to items of self-appraisal. This difficulty, it is believed, is more likely to be found in items of a highly personal nature than in less obvious ones (5).

Adams has made some preliminary study of this problem in connection with Form C of the Personal Audit (1). He administered, using standard instructions, the first seven parts of the form to a class of sophomores (N = 31) in elementary psychology. Three weeks later the same form was administered with these revised instructions:

"You are applying for a job that you want very much. You have been given several tests and have been interviewed by the personnel manager, who tells you that he has an additional test for you to take. You feel that if you pass this test you will get the job which you badly need. Do you understand? Begin."

The two administrations of the form were correlated. It will be noted in Table 3, that while the reliabilities dropped, they are still reasonably high. The usefulness of the form does not seem to be seriously impaired in a simulated employment situation.

TABLE 3

Reliability coefficients between Seven Parts of Form C of the Adams-Lepley Personal Audit Obtained by Retesting
N = 31

Part	Reliability coefficient
I	.69
II	.80
III	.92
IV	.75
V	.75
VI	.69
VII	.63

The nine parts of the present form of the Personal Audit have been tentatively labeled by characteristics which they are thought to measure. It should be emphasized that these names are, at this stage in the development, merely designations which facilitate discussion of the parts of the form. Functional validation which is now in progress will eventually determine what titles can safely be applied.

The various parts of sub-tests of the form may best be described by illustration. Following, under the headings of the tentative labels, are the instructions and the first three statements for each part.

Part I—(Introversion–Extraversion)

Indicate the degree of your liking for each of the following activities by drawing a circle around "M" if you would like it a *great deal,* around "S" if you would have *some* liking for it, around "L" if you would have *little liking* for it, and around "N" if you would have practically *no* liking for it.

1. M S L N Acting as master of ceremonies at a night club.
2. M S L N Introducing strangers at a party.
3. M S L N Playing a part in a movie.

Part II—(Suggestibility)

Fifteen experts, each of whom had won success in a different field, were asked to give an opinion of the statements below. At least eight or more of the experts marked each statement below as *true.* Read each statement, and if you *agree* with the experts that the statement is true, draw a circle around "A." If you agree but only with *reservations,* draw a circle around "R." If you *disagree* with the experts, draw a circle around "D."

1. A R D Persons of widely differing races or faiths should not marry.
2. A R D No further immigration to the United States should be permitted.
3. A R D Only taxpayers should be permitted to vote.

Part III—(Annoyance)

Below is a list of the common annoyances which affect all people to some extent. Indicate your degree of annoyance for each of them by drawing a circle around "M" if it annoys you *much,* around "S" if it annoys you *some,* around "L" if it annoys you a *little,* and around "N" if it *never* annoys you.

1. M S L N To have the doorbell ring while you are taking a bath.
2. M S L N To miss a street car or bus.
3. M S L N To have a stoplight turn red as you drive up to it.

Part IV—(Conflict)

Of the statements below, draw a circle around "T" for the ones you believe to be *usually* true; draw a circle around "D" for the ones whose truth you *doubt;* and draw a circle around "F" for the ones *usually false.*

1. T D F Individuals who are very good-looking usually have less sense than persons with ordinary looks.
2. T D F The girl who uses too much rouge and powder is less modest than the average girl.
3. T D F The scholar who makes a high scholastic average often misses many of the other values of college.

Part V—(Fear)

Below is a list of common fears that all people experience to some extent. Indicate your degree of fear for each of these things by drawing a circle around "M" if you would usually have *considerable* fear, around "S" if you would usually have *some fear,* around "L" if you would usually have *little* fear, and around "N" if you would usually feel *no* fear.

1. M S L N Being trapped in a burning building.
2. M S L N Being buried alive.
3. M S L N Having something fall on you.

Part VI—(Emotionality)

Each word in capitals is followed by four words. Draw a circle around the word that seems to you to go most naturally with the word in capitals.

For example: TRAVEL boat ship train (car)

Here car has been encircled. There are no right or wrong answers. *Work rapidly.*

1.	PAST	yesterday	forget	sorrow	hidden
2.	LEG	limb	walk	shapely	stocking
3.	PURPLE	lavender	royal	wine	crimson

Part VII—(Prejudice)

Indicate the degree of your *dislike* for each of the following activities or things by drawing a circle around

"M" if you would *dislike* it a *great deal,* around "S" if you would *dislike* it some, around "L" if you would have *little dislike* for it, and around "N" if you would have *no dislike* for it

1. M S L N Individuals who bite their nails constantly.
2. M S L N Persons who invariably put on a big "front."
3. M S L N An acquaintance whose breath is offensive.

Part VIII—(Flexibility)

Below is a list of activities or things. If you feel the same way toward them now that you did three or four years ago, draw a circle around "S." If you have *partly* changed your feelings toward them, draw a circle around "P." If your feeling now is considerably *different* from your feelings three or four years ago, draw a circle around "D."

1. S P D Expansion of public employment facilities.
2. S P D Raising the standard of living.
3. S P D Governmental attitude toward labor unions.

Part IX—(Projection)

Below is a list of topics which people consider to some extent at one time or another. Will you indicate the degree of thinking you have given each of them during the *past year* by drawing a circle around "M" if you have done *much* thinking; around "S" if you have done *some* thinking; around "L" if you have done *little* thinking, and around "N" if you have done *no* thinking.

1. M S L N Getting married before the age of 25.
2. M S L N Kissing or petting between young men and women.
3. M S L N Sex attraction as poor reason for marriage.

Scoring of the blank is simplified in that the responses are encircled by the subject, who has merely to choose one in three or four on each subtest. In each part, the responses are arranged in columnar fashion, and each response is given a

TABLE 4

Intercorrelations between Each of the Parts of the Adams-Lepley Personal Audit and the Bernreuter Personality Inventory

(N = 228)

	I	II	III	IV	V	VI	VII	VIII	IX	B1–N	B2–S	B4–D	F1–C
II	−.095												
III	.139	.087											
IV	.374	.109	.204										
V	.067	.222	.591	.360									
VI	.157	.019	.108	.145	.151								
VII	−.067	.226	.418	.194	.393	.094							
VIII	.177	−.036	.113	.126	.234	.047	−.010						
IX	−.033	.332	.225	.174	.280	.079	.331	.153					
B1–N	−.188	.103	.340	.086	.353	.004	.192	.056	.257				
B2–S	−.043	−.092	−.040	−.101	−.418	−.002	−.191	−.218	−.213	−.563			
B4–D	.308	−.124	−.315	.073	−.341	.066	−.179	−.037	−.217	−.861	.632		
F1–C	−.167	.094	.393	.019	.380	−.015	.165	.045	.196	.889	−.624	−.861	
F2–S	−.269	−.031	−.190	−.069	−.254	.011	.001	−.221	−.038	.174	.552	.056	.034

r	P.E.
0 − ± .60	± .04
± .60 − ± .80	± .03
± .80 − ± 1.00	± .02

weight from 0 to 3. One has but to add the responses checked in each column and multiply by the assigned weight; the sums of these products are then the scores of the respective parts.

SOURCE OF DATA

Subjects used in this investigation included 164 summer school students at the Pennsylvania State College (seventy-four male, ninety female), and sixty-four young women who applied for entrance to nursing schools in Clearfield and Philipsburg, Pennsylvania. The summer school students were members of classes in psychology, with the exception of seventeen who were members of a class in engineering economics. Their cooperation was entirely voluntary. The nursing applicants completed the forms as part of a battery of tests used to select new classes of nurses. They were composed of younger girls, and were included to give the sample more heterogeneity. They also provided a basis for determining age differences when later study was made of the data.

TREATMENT OF THE DATA AND RESULTS

Intercorrelations between each of the parts of the Personal Audit and the Personality Inventory were first obtained, using the product-moment method. These are presented in Table 4. N is 228, the whole sample. It will be observed that the vari-

TABLE 5

Parts of the Adams-Lepley Form D Which Are Relatively Independent of Factors in the Bernreuter Personality Inventory and Their Highest Correlation Within the Inventory

Part	Highest r	Trait
I	.308	B4–D
II	−.124	B4–D
IV	.086	B1–N
VI	.066	B4–D
VII	.192	B1–N
VIII	−.221	F2–S
IX	.257	B1–N

ous parts of the Personal Audit are relatively independent of each other. There are some relationships between Part III with Part V and VII, and between Part V and Part VIII, but these are slight.

Of more interest is the independence of the parts of the Personal Audit from the traits of the Personality Inventory. With the exception of Part III and Part V, all are clearly independent factors in terms of the Bernreuter Personality Inventory. Table 5 is set up to illustrate this point.

It is also interesting to note that the highest correlation of any Adams-Lepley trait with a Bernreuter Trait is −.418, between Part V and B2–S.

SEX DIFFERENCES

The means, standard errors of the means, differences between the means, and the ratio of the standard errors to the

TABLE 6

Differences in Mean Scores between Males and Females
(A plus sign indicates males have higher M)

	M_{Male}	M_{Female}	$\sigma_{M_{Male}}$	$\sigma_{M_{Female}}$	D	D/σ_D
Age	26.3	24.3	0.802	0.563	+ 2.0	2.2
I	91.7	85.2	2.45	1.75	+ 6.5	2.2
II	49.2	57.6	2.44	1.26	− 8.4	3.1
III	107.6	112.3	2.67	2.12	− 4.7	1.4
IV	81.2	83.4	2.88	1.84	− 2.2	0.6
V	54.4	71.9	3.07	2.43	−17.5	4.5
VI	28.6	31.2	1.33	1.24	− 2.6	1.4
VII	113.3	127.6	2.91	1.95	−14.3	4.1
VIII	46.2	48.9	2.77	1.97	− 2.7	0.8
IX	59.3	65.5	1.90	1.09	− 6.2	2.8
B1–N	223.8	255.7	8.88	7.14	−31.9	2.8
B2–S	322.3	296.2	6.06	4.22	+26.1	3.5
B4–D	362.5	329.6	6.44	4.96	+32.9	4.0
F1–C	258.4	292.1	9.49	7.91	−33.7	2.7
F2–S	277.9	264.8	6.94	4.20	+13.1	1.6

(A constant, 300, has been added to each Bernreuter Inventory score, to facilitate use of Hollerith machines in handling the data.)

differences, of the male and female segments of the group were studied. Table 6 sets forth these figures. It will be noted that in five instances there are significant sex differences according to the traditional concept of the standard error ratio (t). However, if we follow the suggestion of Peters (7), a standard error ratio of 1.73 will be regarded as an indication of an acceptable reliable difference, and five additional t's may be regarded as reliable.

Within the limits of our terminology this may mean that, in general, women are more liable to suggestion than men (Part II); that women have considerably more fears, or fear capacity, than men (Part V); that in general women are more prejudiced than men (Part VII); that men are more self-sufficient than women (B2–S); and women are generally more submissive than men; that men tend to be more "extraverted" than women (Part I); that women rationalize more than men (Part IX); that women are inclined to be more neurotic than men (B1–N); and that men are more self-confident than women (F1–C).

Since it was convenient to do so, the same comparisons were made between the male group and the summer school female group within the sample. This eliminated the younger applicants. In view of the fact that of the ninety women in this comparison sixty-six are teachers and seven are public school nurses, composing 81 per cent of the group, it is interesting to note several of the *t*'s. On Part II, *t* is smaller (1.9) when the more mature group of women is compared with the men. This would seem to indicate that as women grow older they show less tendency to be suggestible. On Part V, *t* is also smaller (3.6), which shows some tendency for women to forget, or to adapt to their fears somewhat. Increases in the size of *t* are noted on Parts III (2.1), B1–N (4.6), B2–S (4.3), B4–D (5.1), and F1–C (4.7), indicating a tendency on the part of older women to become more easily annoyed than men, more unstable than men, less self-sufficient than men, more submissive than men, and that they come to be less confident in them-

selves than men. It may be that this latter comparison shows traits more characteristic of teachers than of women in general. Because the population used in this study is not wholly made up of teachers no such generalization can be made. On the other hand, since the population in Table 6 is admittedly more heterogeneous, and since there are corresponding reliable differences among the Parts, it is possible to generalize from this table with caution.

TABLE 7

Differences in Mean Scores between Younger and Older Females
(A plus sign indicates younger group has higher mean score)

	$M_{Younger}$	M_{Older}	$\sigma_{M_{Younger}}$	$\sigma_{M_{Older}}$	D	D/σ_d
Age	19.2	28.0	0.28	0.72	− 8.8	11.4
I	87.6	83.4	2.51	2.45	+ 4.2	1.2
II	61.5	54.9	1.61	1.76	+ 6.6	2.8
III	108.2	115.3	3.69	2.44	− 7.1	1.6
IV	82.7	83.9	3.09	2.27	− 1.2	0.3
V	74.8	69.9	3.94	3.05	+ 4.9	1.0
VI	31.5	31.0	1.75	1.71	+ 0.5	0.2
VII	127.3	127.9	3.26	2.37	− 0.6	0.1
VIII	56.7	43.4	3.21	2.32	+13.3	3.4
IX	67.5	64.0	1.89	1.33	+ 3.5	1.5
B1–N	216.2	283.8	8.55	9.59	−67.6	5.3
B2–S	310.6	286.0	5.64	5.77	+24.6	3.0
B4–D	350.3	314.8	6.60	6.83	+35.5	3.7
F1–C	245.8	325.0	9.75	10.35	−79.2	5.6
F2–S	262.1	266.7	5.81	5.89	− 4.6	0.6

(A constant, 300, has been added to each Bernreuter Inventory score, to facilitate use of Hollerith machines in handling the data.)

AGE DIFFERENCES BETWEEN TWO GROUPS OF THE SAME SEX

Inasmuch as there were two rather distinct age groups among the female population of the sample, they were compared in order to find the effect of age upon the scores obtained. In Table 7 are found the differences in mean scores between the younger women (nursing applicants) and the summer school women.

The standard error ratio is significant in Parts II, VIII, B1–N, B2–S, B4–D, and F1–C. This indicates a tendency on the part of the older women to be less moved by suggestion than the younger women (Part II); the younger women tend to be more flexible in their thinking and attitudes (Part VIII); the older women tend to be more neurotic than younger women (B1–N); the older women are more self-sufficient than the younger (B2–S); the older women tend to be more dominant in face-to-face situations than the younger (B4–D); and the younger women seem to be more self-confident (F1–C).

CONCLUSIONS

From this study of interrelationships between the Adams-Lepley Personal Audit and the Bernreuter Personality Inventory, the following conclusions might be drawn.

1. Intercorrelations between the parts of the Personal Audit have been obtained which are substantially the same as those reported by the authors of the form.

2. Intercorrelations between the parts of the Bernreuter Personality Inventory have been obtained which are substantially the same as those reported by the author in the Manual for the Personality Inventory.

3. Correlations between the parts of Form D with the Bernreuter factors reveal that Parts I, II, IV, VI, VII, VIII, and IX of Form D seem to be quite independent of the Bernreuter factors.

4. Reliable sex differences were found to occur in all but the factor F2–S of the Personality Inventory.

5. Reliable sex differences were found to occur in all but Parts IV, VI, and VIII of Form D.

6. Reliable differences due to age were found among females in each of the Personality Inventory Factors with the exception of F2–S.

7. Reliable differences due to age were found among females in Parts II and VIII of Form D.

BIBLIOGRAPHY

1. ADAMS, C. R. *Personality Test Results in a Simulated Situation.* Unpublished study.
2. BERNREUTER, R. G. *The Evaluation of a Proposed New Method for Constructing Personality Trait Tests.* Stanford University, California: Stanford University Library, 1931.
3. ————————. The imbrication of tests of introversion–extroversion and neurotic tendency. *Psychol. Bull.*, 1933, 30, 665–666.
4. ————————. *Manual for the Personality Inventory.* Stanford University, California: Stanford University Press, 1935.
5. ————————. *The present status of personality trait tests. Ed. Rec.*, 1940, 21, 160–171.
6. FLANAGAN, J. C. *Factor Analysis in the Study of Personality.* Stanford University Press, 1935.
7. PETERS, C. C., AND VANVOORHIS, W. R. *Statistical Procedures and Their Mathematical Bases.* New York: McGraw-Hill, 1940.
8. SLETTO, R. F. *Construction of Personality Scales by the Criterion of Internal Consistency.* The Sociological Press, 1937.

THE VALIDITY AND RELIABILITY OF THE VAN RIPER CRITICAL ANGLE BOARD IN DIAGNOSING HANDEDNESS

W. E. COOPER

Principal of Stonewall Jackson School, Denton, Texas

AND

M. E. BONNEY

North Texas State Teachers College, Denton, Texas

THE interest of students and teachers in the phenomenon of human handedness is by no means new. A large number of investigations have been carried on with the aim of answering various questions which are raised by the relation of man's hands to each other. A wide variety of attacks have been made on the problem. This study is concerned with the validity and reliability of the Van Riper Critical Angle Board in diagnosing handedness.

A survey of one hundred twenty-five elementary pupils in the schools of Denton, Texas, with the Van Riper test made possible the selection of a small group for more intensive study. Thirty definitely right-handed children were chosen. Out of the 125 tested only 8 were clearly left-handed. This corresponds closely with the general norm for the extent of left-handedness which has been established by other studies such as that by Gordon in which he found 7.3 per cent of a total of 3,298 children of school age to be left-handed (4). In order to have thirty left-handed cases to match the thirty right-handed cases a number of teachers were asked to select pupils whom they thought to be predominantly left-handed. A total of 30 was secured in this way.

Each child's performance on the critical angle board was checked against his hand preference in three other test situ-

ations, namely, throwing a dart at a target, pitching bean bags, and putting pegs in a peg-board. This testing was done in a separate room where the subjects were not disturbed by school activities. One child was tested at a time. In none of the tests did the subjects know that their handedness was being observed. In each of the tests the child was allowed to perform with whichever hand he preferred, and after completing the series he was told that this was to be a race between the hands. He was to see if he could do better with the other hand. His attention was focused on the scores being kept as the main object of the games. This was done to prevent the factor of self-consciousness from influencing hand preference.

A brief description of the equipment used in the performance tests follows.

The dart test equipment consisted of eight ordinary darts and a target board, two feet square, with three circles painted on it. The center circle was marked 15, the next circle 10, and the last circle was marked 5 for convenience in scoring. The target board was hung on the wall about five feet from the subject. The darts were placed on a table. The examiner explained to the subject that he was to throw the darts at the board and to try for the highest score possible. The hand volunteered was recorded as the preferred hand. A record was kept of the score made on each of the eight throws. These were later compared with accuracy in throwing with the non-preferred hand.

Ten four-inch bean bags and a paper box were the required materials used for the bean bag test. The box was twelve inches square and six inches deep with a circular hole in the top five inches in diameter. The subject was instructed to throw the bean bags in the box from a distance of five feet. Every time a bag fell in the box a score of 10 was given. If the bag fell on the top of the box a score of 5 was given, and if it fell on the floor no score was given. When the subject picked up the first bag the hand volunteered was recorded. He was allowed to continue throwing the bags with the pre-

ferred hand. The scores were recorded and kept for comparison with the non-preferred hand.

The peg board was nine inches square with small holes every one-fourth inch for the pegs. A stop watch was used to time each subject for one minute. The subject was instructed to begin putting the pegs in the holes when the signal "go" was given and to continue until the word "stop." The hand volunteered was recorded and the number of pegs placed in the holes made the score.

Other investigators have used tests identical or similar to those just described to determine handedness. Dart (2) found pointing and reaching to be the most reliable. Turner (12) found throwing to be the best since it showed only 3.5 per cent inconsistency, that is, only 3.5 per cent did better with the non-preferred hand. In Turner's study there was a 5.3 per cent inconsistency for a pointing test and 9.5 per cent in energetic reaching. Haefner (5) states that if one can observe the throwing activity of an individual, he can obtain an idea as to dominant handedness which is correct in about nine out of ten cases for both left and right-handed subjects. He also states that easy reaching would be an index in eight out of ten cases, and energetic reaching in seven or eight out of ten cases. In Beeley's (1) study a very close agreement between the tracing board with the stylus and known handedness was found. Ojemann (8), Koch (6), and Haefner (5) emphasize that a battery of tests is always more reliable than any one test. The present study bears out this emphasis in that it was found that none of the three uni-manual tests upon which scores were obtained was as satisfactory a measure of handedness as a combined score. For example, the amount of inconsistency in the bean bag throwing test was 13 per cent, that is, 8 children out of 60 (13 per cent) made a better score with the non-preferred hand, whereas the average inconsistency for the three performance tests was only 8 per cent. Nearly all of the subjects who made a better score with the non-preferred hand were left-handed children who did better with the right hand. This is

probably due to left-handed children being encouraged by parents and teachers to use their right hands for some things whereas right-handed children are seldom, if ever, encouraged to use their left hands.

In administering the Van Riper Critical Angle Board there are three patterns for the subjects to draw, namely, the kinesthetic, the script, and the visual.

The kinesthetic pattern used was the profile of the body of a bear. This was grooved in a board one inch thick, twelve inches long, and six inches wide. The subject was blindfolded and a stylus was placed in his hands. The blindfold is recommended by Van Riper (13) to develop a muscular perception of the pattern for reproduction on the vertical writing board. The subject was instructed to trace the grooved pattern until he was sure that he knew it well enough to draw it. He was allowed to draw the pattern on the table top, and if not successful he was allowed to do the tracing exercise again. At this point, the subject was seated before the critical angle board and directed to draw with each hand the pattern just memorized. It was drawn simultaneously on each wing of the angle board. The wing boards were moved ten degrees away from the subject after each complete pattern. One trial was given beyond the mirrored (backward drawing) pattern as a check. The degrees at which mirroring occurred was recorded as the score. The hand mirroring was designated with "R" or an "L" as the case might be. On the Van Riper test the normal expectancy is that the right handed subjects, when they reach the critical angle for them, will draw the patterns backwards with the left hand, and that the left-handed subjects likewise will mirror with the right hand. The subjects are not conscious that they are making the patterns backwards.

In administering the script pattern, each subject was blindfolded and given two pencils and as the examiner pronounced the word "cat" he was instructed to write it as quickly as possible. The wings of the angle board were moved ten degrees after each complete pattern until mirroring. One trial was given beyond the first mirrored pattern.

The visual pattern was in this form . This was hung on the wall in front of the subject. A cardboard was held between the subject's eyes and hands to eliminate any observation of hand movements while reproducing the pattern. As before, the wings of the angle board were moved ten degrees away from the subject after each drawing and one trial was given beyond the first mirrored pattern as a check.

Figure 1 is a diagram of the Van Riper Critical Angle Board.

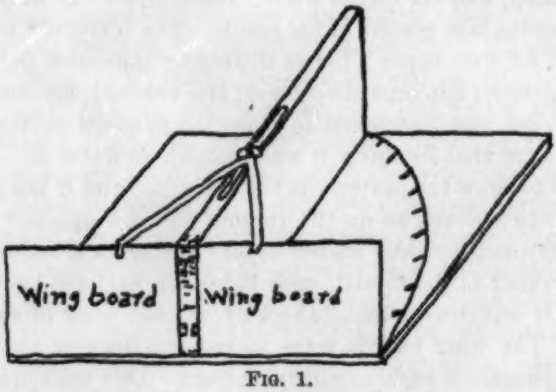


FIG. 1.

The critical angle at which mirroring occurs on the Van Riper test varies from subject to subject according to the degree of right or left-handedness. Also there is considerable variation for the three types of patterns. The average critical angles for left-handed subjects in this study were 35, 58, and 36 degrees for the kinesthetic, script, and visual patterns, respectively. The average angles for the right-handed subjects were 53, 49, and 49 degrees for the kinesthetic, script, and visual patterns.

In order to check the validity of the critical angle board against the criterion of handedness as determined in the performance tests two total sets of scores were necessary. A composite score for each child on the performance tests was obtained for each hand by determining the average score made in

throwing the 8 darts, the 10 bean bags, and in placing the pegs in the peg board during the one minute allowed. Likewise a total score on the Van Riper test was secured by taking an average of the degrees at which each child mirrored on the three patterns described above. Using the Pearson Product Moment method a coefficient of correlation of $+.98 \pm .004$ was obtained between these two sets of scores for the 30 left-handed pupils and a coefficient of $+.93 \pm .015$ for the 30 right-handed children. Such a degree of relationship certainly justifies the conclusion that the Van Riper Critical Angle Board is a valid method of measuring handedness particularly in subjects who are definitely right or left-handed.

The reliability of the critical angle board was determined on the Denton subjects by repeating the Van Riper test approximately a year after the first tests were given (December, 1940). Of the original 60 children, twenty-seven right-handed and nineteen left-handed ones were available. The correlation for the two tests on the right-handed children was $+.90 \pm .02$ and for the left-handed it was $+.98 \pm .02$. Coefficients of such magnitude are certainly high enough to warrant the conclusion that the critical angle board is a reliable test of handedness in definitely right and left-handed children. These findings are in accord with results published by Van Riper (13).

In order to get some data on the value of the critical angle board in diagnosing handedness with adults, thirty students from the North Texas State Teachers College were asked to take the Van Riper test. After each one had finished he was told his predominate handedness and this was checked against the student's experience. In no case was the diagnosis in error.

From the above data it would seem that the critical angle board should be given serious consideration by those engaged in clinical practice in which a quick measure of handedness is desired.

BIBLIOGRAPHY

1. Beeley, A. L. An experimental study in left-handedness. *Supplementary Educational Monograph*, II, No. 2.

2. DART, C. E. Hand, eye, and foot preferences of 200 mentally subnormal subjects compared with that of subjects of normal intelligence or above. *J. Juvenile Res.*, April, 1938, XXII, No. 2, 119–121.
3. DOWNEY, J. E. Laterality of function. *Psychol. Bull.*, 1933, XXX, 108–142.
4. GORDON, H. Left-handedness and mirror writing. *Brain*, 1920, XLIII, 530–538.
5. HAEFNER, RALPH. *The Educational Significance of Left-Handedness.* Teachers College Contributions to Education, 1929, No. 360.
6. KOCH, H. Study of the nature, measurement and determination of hand preference. *Genet. Psychol. Monog.*, 1933, XIII, 117–218.
7. MEYER, M. Left-handedness and right-handedness in infancy. *Psychol. Bull.*, 1913, X, 52–53.
8. OJEMANN, R. H. Studies in handedness: I. A technique for testing uni-manual handedness. *J. Educ. Psychol.*, 1930, XXI, 597–612.
9. SELZER, C. A. *Lateral Dominance and Visual Fusion.* Cambridge: Harvard University Press, 1933.
10. STRANGER, M. A., AND DONOHUE, E. K. *Prediction and Prevention of Reading Difficulties.* New York: Oxford University, 1937.
11. TRAVIS, L. E. Recurrence of stuttering following shift from normal to mirror writing. *Arch. Neurol. and Psychiat.*, 1929, XXI, 386–391.
12. TURNER, E. Studies in eye, hand, and foot preferences of emotionally unstable adolescents compared with stable adolescents. *J. Juvenile Res.*, April, 1939, XXII, No. 2, 122–124.
13. VAN RIPER, C. A new test of laterality. *J. Exper. Psychol.*, April, 1934, XVIII, 295–313.
14. ————. The quantitative measurement of laterality. *J. Exper. Psychol.*, June, 1935, XVIII, 372–382.

THE EFFECT OF CHANGES IN VITAMIN A CONTENT IN THE DIET UPON RECOVERY FROM GLARE BLINDNESS

C. W. BROWN, L. B. FISK, AND H. P. TORKELSON
University of California

IN recent years there has been considerable interest manifest in the relationship between vitamin A intake and glare blindness, sometimes called night blindness. It is a familiar fact that when the eye fixates under conditions of low illumination, after being stimulated with a bright light, the resulting experience is one of total blackness, no differentiation of visual objects being possible. Then the eye gradually recovers and differentiation of objects in the dark becomes increasingly better. Glare blindness refers to the rate of recovery of the eyes after stimulation with a bright light. The individual who recovers very slowly is considered to be glare weak. A quantitative measure of the recovery process is obtained by first stimulating the eyes with a bright light and then noting the change in the brightness threshold during a subsequent continuous exposure to the dark, the value of the threshold decreasing as recovery takes place.

The association of this characteristic of the light sense with malnutrition, particularly with deficiency in vitamin A, has been known for many years (Fredericia and Holm, 1925). Tansley (1933) has shown that the concentration of visual purple in the retinas of rats and dogs, which have been subjected to a vitamin A deficient diet, is very low and may be totally absent. She (1931) previously had presented evidence which indicated that vitamin A is a precursor of visual purple, and that the amount of visual purple in the eye at any given

time is in part conditioned upon the amount of vitamin A present. Some investigators have used the relationship between the rate of recovery following stimulation with a bright light and vitamin A as a basis for detecting subclinical states of vitamin A deficiency (Jeans and Zentmire, 1934; Park, 1936; Jeans, Blanchard, and Zentmire, 1937; Jeghers, 1937; Hecht and Mandelbaum, 1939; Wald and Stevens, 1939). In the report of Jeans and Zentmire (1934), 21% of a group of 213 children were considered to be suffering from avitaminosis A. On testing 275 adults, Park (1936) found two in a group considered to be vitamin A deficient who also experienced difficulties when driving at night due to the glare of the headlights of approaching cars. Some of Jeghers' (1937) subjects also experienced similar difficulties. On questioning those individuals who had long recovery times, several of them reported that they preferred not to do any night driving because of experiencing relatively long periods of blindness from the headlights of approaching cars. In a later and more extensive study (1938) those subjects having difficulties in night driving were given large doses of vitamin A concentrates. They all gained relief and could again drive at night with little ill effect from glaring headlights. In these investigations the photometer or some adaptation of it was used, and the measure of the dark adaptation rate was the rate of change in the threshold for absolute intensity following stimulation with a very bright light.

The present writers became interested in the relationship between glare weakness and vitamin A intake in connection with studies on the factors conditioning efficiency in automobile driving. Investigations carried out in conjunction with the California Department of Motor Vehicles had shown that there is a general tendency for good drivers to have better resistance to glare than poor drivers.[1] It occurred to the

[1] These studies were made in conjunction with Mr. E. D. Fletcher of the Department of Motor Vehicles. The glare resistance apparatus used was an improvement of the De Silva glare test.

writers, that with the close association known to exist between vitamin A and visual purple, it might be possible to show a relationship between vitamin A content in diet and recovery time from glare stimulation. The demonstration of this relationship would be an important step in the attack upon the problem of night accidents.

That the rate of dark adaptation can be altered by changes in the vitamin A intake has been demonstrated by other investigators (Jeghers, 1938; Hecht and Mandelbaum, 1939; Wald and Stevens, 1939). The present study differs from these investigations in several respects. In the first place, the method of measuring the eyes' recovery from light stimulation by means of the photometer or some adaptation of it is too complex and requires too much time for this instrument to be used on the wide scale testing demanded in the examining of hundreds of thousands of automobile drivers. Consequently, a similar method, requiring only a few minutes of testing time, is needed. Secondly, changes in the vitamin A intake through changes in the diet of the subjects, as contrasted with administering vitamin A concentrates, needed to be studied. It was thought that any wide use of vitamin A therapy in the treatment of automobile drivers suffering from glare weakness could be best accomplished through correct regulation of dietary requirements. Third, the factor of individual variations, which had been a disturbing factor in most of the preceding studies, merited a more careful evaluation.

THE GLARE RECOVERY APPARATUS

The apparatus[2] for measuring recovery from glare stimulation consisted of two parts, one to control the presentation of a bright light, the other to present a target under low illumination and measure the time of recovery. The bright light was produced by adapting parts of the headlights of an automobile. Two lenses were fastened together with the ripples in

[2] A detailed description of the apparatus is given in an article by Brown in the JOURNAL OF GENERAL PSYCHOLOGY. 1941, 25, 421–428.

the glass at right angles in order to give the greatest diffusion of light. The lenses were then placed at the correct focal distance from a reflector in which was mounted a six-volt light globe. All of these parts were mounted in a red wood box, measuring 20″ long, 10″ wide, and 9″ high. At the end of the box facing the lenses, a visor was arranged through which the subject could view the light. Immediately in front of the lenses was placed a small white cardboard arrow, which could be controlled from the outside, and which could be rotated to any position at right angles to the subject's line of regard. A small light was provided to illuminate the target when the bright light was not turned on. Ordinary 110-volt house current was adapted to the needs of the apparatus by means of a transformer. A rheostat controlled the brilliance of the bright light. The time of recovery was measured by means of a stop watch.

METHOD OF MEASURING THE RECOVERY FROM GLARE STIMULATION[3]

The subject was seated in front of the apparatus with his head held tightly against the visor. With the dim light on, he was shown how the arrow would be rotated to different positions and instructed that he was to report as soon as he could the exact position of the arrow. He also was informed of the brightness of the preadapting stimulating light, and the exact time he would be exposed to it. The bright light then was turned on, and gradually increased in brilliance to the maximum through a period of 20 seconds, and then left on at the maximum for another period of 20 seconds. The bright light then was turned off, and the dim light turned on. At the same time the stop watch was started. The subject reported as soon as he could see the arrow well enough to tell the direction in which it was pointing. The arrow was then changed in position several times and at least three responses

[3] Full details on the method of measuring the recovery from glare stimulation will be found in the article in the JOURNAL OF GENERAL PSYCHOLOGY.

obtained from the subject. This was done in order to make sure that the first recognition was not a chance response. The watch was stopped after the third correct response. The measure of the recovery time was the number of seconds required by the subject to recover sufficiently from the glare stimulation that he could correctly identify the position of the arrow presented under relatively low illumination. Although this index does not measure changes in the absolute intensity threshold during dark adaptation, it does give an accurate estimate of visual efficiency under conditions of glare stimulation.

PROCEDURE FOR VARYING AND MEASURING THE VITAMIN A INTAKE

Differences in the vitamin A intake were achieved by having the subjects follow diets low and high in vitamin A content. Charts were prepared, based on Daniel and Munsell's "Vitamin Content of Foods" (1937), which gave the estimated vitamin A content of 85 common foods. During one phase of the experiment the subjects confined their eating as much as possible to foods rich in vitamin A, and during another phase to foods low in vitamin A.

Daily records were kept by the subject in which he indicated the foods eaten and the size of the helpings consumed. Helpings were recorded as small, medium, or large. At the completion of the experiment the amount of the vitamin expressed in Shermund units in each kind of food for the three sizes of helpings was estimated, and the daily intake of vitamin A computed. Weighing each helping would have been more accurate, but it was wholly impracticable. The method used assumes that an individual's estimate of the size of the helpings of foods he consumes will not vary greatly between the three diet conditions. Measurement of the vitamin intake seems to have been sufficiently accurate for the purposes of the present study.

GENERAL CONDUCT OF THE EXPERIMENT

Ten college students participated in the study, seven being seniors and three graduate students. For several days pre-

ceding the introduction of a special dietary regimen, records were made of the food intake and of the scores on the glare recovery apparatus under what can be considered normal diet conditions. These measurements constituted a means for evaluating the recovery responses taken under the two special diet conditions. The subjects then were divided randomly into two groups. Group one was placed on a diet low in vitamin A for a period of 12 days, and then was transferred to a diet high in the vitamin for another 12 days. Group two was given the high diet first and then was transferred to the low diet. Records of the recovery from glare stimulation were made every day during the week and on Saturdays and Sundays whenever possible. On Tuesday and Thursday of each week two recovery-from-glare scores were obtained in the afternoon, separated by an interval of from one to two hours.

RELIABILITY OF THE GLARE RECOVERY APPARATUS

Two checks were made on the consistency of measurement of the apparatus. In the first study the recovery scores were taken on 86 college students at two different times, with several days intervening between the tests. If the subject had been working under a bright light source or had come in from bright sunlight conditions, the test was delayed for a few minutes until his eyes had adjusted to the illumination of the experimental room. The Pearsonian correlation coefficient between the scores on the two tests was .77. Although this is not as high as is usually required for individual diagnosis, it does show a creditable consistency when the very short time involved in testing (usually less than 2 minutes) is taken into consideration.

The present experiment afforded another rough check on the reliability of the apparatus. The pairs of scores obtained on Tuesday and Thursday of each week were correlated, giving a coefficient of .81. Inasmuch as each score of a pair of scores was obtained on the same day and at about the same time of day, individual differences resulting from diurnal variations

and variations from day to day were minimized. However, dietary changes were involved during the period in which these scores were taken, and these changes would tend to increase the variation in the individuals' scores. As a consequence the variability of the scores in the diet experiment with ten subjects was not significantly smaller than the variation in the scores obtained in the non-diet experiment with 86 college students. The available evidence supports the contention that the true reliability of the apparatus is about .80, which is sufficiently high for studying group differences and for detecting severe weaknesses in glare recovery.

IMMEDIACY OF THE EFFECTS FROM CHANGING TO A DEFICIENT OR TO AN EXCESS VITAMIN A DIET

Due to the small number of cases studied it is impossible to determine just how quickly the ability to recover from glare stimulation is affected by changes in the diet. In Group I, which was given the deficient diet first, the recovery times increased during the first five days of the diet, after which they remained relatively the same, at a high level, during the remainder of this diet condition. When Group II was changed from an excess to a deficient diet, most of the increase in the recovery scores was reflected in the responses of the first day after the change. Group II had been consuming an adequate amount of vitamin A in the normal diet condition and therefore showed no decrease in score when placed on the excess diet. On the other hand, when Group I was transferred from a deficient to an excess diet, their recovery scores gradually decreased for four days, after which there was no significant change. The foregoing statements are to be accepted as only suggestive of the times required for changes in vitamin A intake to produce measurable differences on the recovery-from-glare scores.

ANALYSIS OF VARIANCE

By means of the statistical methods involving variance it is possible to separate the variation in the scores that resulted

from differences within the individual subjects from the variation attributable to dietary changes. In Table 1 an analysis

TABLE 1

Analysis of Variance Attributable to Diet Changes, Individual Differences and Error Factors

Source	Degrees of freedom	Sum of squares	Variance	F	1% level
Diet Changes	2	329.52	164.6	7.18	6.01
Individual differences	9	764.06	83.9	3.66	3.60
Error factors	18	412.26	22.9		
Total	29	1505.84			

of variance is given, following Fisher's method (Fisher, 1936). There are three variations in the dietary treatment: normal diet, high vitamin A intake, and low vitamin A intake. There are then two degrees of freedom for the diet factors. There are records from ten subjects so the degrees of freedom for the individual difference factors are nine in number. The total degrees of freedom as the analysis is set up are 29, or one less than the total number of values entering into the analysis. The difference between the total degrees of freedom and the degrees of freedom accounted for in the diet and individual difference factors is 18, which are the degrees of freedom to be assigned to error factors.

The sums of the squares for the different factors are given in column two; for the diet differences this sum is 329.52, for individual differences 764.06, and for error factors 412.26. In the third column the variances are given as follows: 164.6 for the diet differences, 83.9 for the individual differences, and 22.9 for the error factors. The significance of the variance resulting from diet changes can be determined by using Fisher's constant F. The value of F for the diet changes is 7.18. For the variance of the diets to be significant at the 1% level, F should have a value of at least 6.01. It is apparent

from the table that the value of F exceeds the 1% level, and therefore, changes in vitamin A intake produce significant changes in the time of recovery from glare stimulation.

DIFFERENCES IN THE RECOVERY SCORES UNDER THE DEFICIENT AND EXCESS VITAMIN A CONDITIONS

A further analysis of the results are given in Table 2. The

TABLE 2

Significance of Difference in Recovery Times between Excess and Deficient Vitamin A Diet Conditions

Source of measures	Time of recovery		Difference	S.E. of difference	Value of t
	Deficient diet	Excess diet			
Averages from all days	22.70	18.03	4.67	1.28	3.65
Scores from last day only	24.25	15.25	9.00	3.21	2.80

mean times of recovery for the ten subjects under the conditions of high and low vitamin A intake are presented in row one. The mean time for the deficient vitamin A condition is 22.70 seconds, and for the excess vitamin A condition 18.03. The difference is 4.67 seconds, with a standard error of 1.28. The value of t (see Fisher, 1936) for this difference is 3.65. The value of t at the 1% level of significance for nine degrees of freedom is 3.25. The difference then is highly significant and it can be concluded that recovery from glare stimulation is significantly affected by changes in the vitamin A intake in the diet.

To determine if this difference, found for the averages for the entire experimental period, would also occur for a single day's responses, the recovery times recorded for the last day under each dietary regimen are compared. These recovery times are presented in the second row of the table. The mean values are: for the low vitamin A intake, 24.25 seconds, and

for the high, 15.25 seconds. It will be observed that the difference of 9.00 seconds is almost twice as large as was obtained between the former averages. However, it also will be observed that the standard error of the difference is more than doubled. Nevertheless, the difference is highly significant. The value of t is 2.80 which is slightly less than the value required at the 2% level of significance. These findings clearly demonstrate that radical changes in vitamin A intake affect the recovery time from glare stimulation.

DEGREE OF RELATIONSHIP BETWEEN VITAMIN A INTAKE AND RECOVERY FROM GLARE STIMULATION

A rough approximation of the degree of relationship existing between the two variables, vitamin A intake and recovery from glare, can be obtained by calculating the rank order correlations between the amount of vitamin A in the diet and the recovery time after glare stimulation, using the mean scores of the ten subjects. The coefficient is −.41 as given in the first row of Table 3. This coefficient is not significant for

TABLE 3

Degree of Relationship between Amount of Vitamin A in Diet and Recovery Time

Rho between averages N equals 10	−.41				
Rho for each subject N 12 to 21	−.78	−.72	−.08	−.26	−.70
	−.48	−.57	−.13	−.48	+.13
Average of ten coefficients	−.46				

an N of only 10 cases. However, the relationship is substantiated when a rank coefficient is computed for each of the ten subjects. These appear in the second row of the table. Six of the 10 coefficients are significant at the 5% level. They are underlined in the table, and are −.78, −.72, −.70, −.48, −.57, and −.48. Three of the remaining four coefficients are in the

right direction but are not statistically significant. One coefficient is positive but insignificant. The subject who furnished the data for this coefficient showed no difference in his recovery time for the two dietary conditions. The reason for certain individuals' responding in this manner is still unknown. Wald and Stevens (1939) report that in one of their experiments, half of the subjects showed no sensitivity to glare.

If it can be assumed that these rho coefficients are approximately equivalent to Pearsonian r's then it is possible to get a coefficient which is an estimate of the average relationship among the two variables. Using Fisher's method, involving the z-equivalents of r (Fisher, 1936), the average correlation comes out to be $-.46$. This is approximately the same value as the rho calculated between the mean scores of the ten subjects, which was $-.41$. These coefficients indicate that under the conditions of the present experiment there is a negative and moderately close correlation between the amount of vitamin A in the diet and the time of recovery from glare stimulation.

SUMMARY

A simple apparatus for measuring recovery from glare stimulation, which can be used in the examination of automobile drivers, is described. The apparatus is found to be sufficiently reliable to be used for measuring group differences and for diagnosing severe weakness in glare recovery. Variations due to radical changes in the vitamin A intake in the diet are shown to cause changes in the recovery score. Subjects on a diet low in vitamin A intake have a significantly larger recovery score than they have when they are on a diet high in vitamin A. A moderate, but substantial correlation is found between the amount of vitamin A consumed and the time required to recover from glare stimulation.

BIBLIOGRAPHY

DANIEL, E. P., AND MUNSELL, H. E. *Vitamin Content of Food.* Washington: U. S. Dept. Agric. Misc. Pub. No. 275, June, 1937.

FISHER, R. A. *Statistical Methods for Research Workers.* Edinburgh: Oliver & Boyd, 1936.

FREDERICIA, L. A., AND HOLM, E. Experimental contribution to the study of the relation between night blindness and malnutrition: Influence of deficiency of fat soluble A-vitamin in the diet on the visual purple in the eyes of rats. *Am. J. Physiol.*, 1925, 73, 63–78.

HECHT, S., AND MANDELBAUM, D. Dark adaptation and vitamin A. *J. Am. Med. Assoc.*, 1939, 112, 2, 1910–1916.

JEANS, P. C., BLANCHARD, E., AND ZENTMIRE, Z. Dark adaptation and vitamin A—A new photometric technique. *J. Am. Med. Assoc.*, 1937, 108, 451–458.

JEANS, P. C., AND ZENTMIRE, Z. A clinical method for determining moderate degrees of vitamin A deficiency. *J. Am. Med. Assoc.*, 1934, 102, 892–895.

JEGHERS, H. Night blindness due to vitamin A deficiency: Consideration of its importance in traffic problems. *New England J. Med.*, 1937, 216, 51–60.

———. Vitamin A deficiency in adults. *Internat. Clin.*, 1938, 1, 17–33.

PARK, I. O. Further observations on vitamin A deficiency as shown by studies with visual photometer and clinically. *J. Oklahoma Med. Assoc.*, 1936, 29, 129–135.

TANSLEY, KATHERINE. The regeneration of visual purple; its relation to dark adaptation and night blindness. *J. Physiol.*, 1931, 71, 442–458.

———. Factors affecting the development and regeneration of visual purple in mammalian retina. *Proc. Roy. Soc. London*, Series B, 1933, 114, 79–102.

WALD, G., AND STEVENS, D. Experiment in human vitamin A deficiency. *Proc. Nat. Acad. Sci.*, 1939, 25, 344–349.

AUDIENCE TENDENCIES IN RATING PUBLIC SPEAKERS

ALICE I. BRYAN
Columbia University
AND
WALTER H. WILKE
New York University

A TECHNIQUE for measuring the reaction of an audience to a public speech[1] has recently been described. Published data on the B–W Scale for Rating Public Speeches include information regarding reliability and validity of the technique, percentile norms, and representation of audience reactions to a speaker in the form of a speech profile (1, 2). The present article, based on data obtained through use of the scale by a variety of audiences, is concerned with some of the factors associated with audience ratings.

ANALYTICAL ABILITY OF AUDIENCE

Use of a rating scale by members of an audience assumes some degree of analytical ability. The listener is expected to indicate his reaction to such varied aspects of delivery and subject matter of a speech as clarity, appearance and personality of speaker, and command of language. Doubt has been thrown on the ability of an audience to discriminate between items of this type by Monroe, Remmers, and Lyle (3), who studied the specificity of ratings on the Purdue rating chart. They found average inter-correlations between three items (enthusiasm, attention, and organization) of the order of .60 for ratings by the same individuals and average r's of the order of .30 for ratings of a given item by different individuals.

[1] *The Bryan-Wilke Scale for Rating Public Speeches*, revised edition. The Psychological Corporation, New York, 1939.

Since the reliability of the rating was of the order of .30, it was concluded that a rating for any one of these three traits was equally good as a measurement of either of the others.

It seemed likely that some of the items on the Purdue chart (directness, objectivity, motivation, etc.) would prove more difficult for the average member of an audience to define and somewhat more likely to overlap in meaning than the items of the B–W Scale. An analysis of the relation between separate items of the latter scale was therefore undertaken. It had already been found that one group of items correlated more closely than another group (1, Table II, p. 89) when seven pairs of items considered likely to include a minimum of common factors were contrasted with seven other pairs which appeared more obviously dependent upon common elements. The correlations obtained were in accord with differences to be anticipated if the ratings were reasonably valid. This finding indicates that to some degree different items of the scale measure different reactions.

A more direct measure of the effect of the listener's general impression as manifested by ratings on separate items of the scale would be indicated by the correlation between ratings on single items and ratings on the scale as a whole, the latter taken as a measure of general impression. The correlations presented in Table 1 are based on the average of 20 ratings on each of 36 speakers.

Some correlation between individual items and the scale as a whole would be expected for several reasons, such as the influence of previous experience in speaking, etc., but correlations as high as those indicated in Table 1 are probably due in considerable measure to halo effect or general impression. Partial correlations, holding constant the total score or general impression, were therefore computed between each of the fourteen pairs of items previously referred to (1, Table II, p. 89). These partial correlations are given in Table 2. Although these figures are based on over 500 ratings, the number of speakers involved is too small to warrant attaching much sig-

TABLE 1

Correlation between Total Scores and Individual Items

Item	r	Item	r
1	.548	9	.745
2	.718	10	.702
3	.815	11	.776
4	.839	12	.689
5	.911	13	.839
6	.812	14	.697
7	.893	15	.620
8	.638	16	.449

nificance to any one inter-correlation. The general trend is so clear, however, that certain facts are probably characteristic of such audience ratings.

The fourteen inter-correlations listed—chosen for study from the 120 possible r's because they presumably represented the extremes of high and low relationships—show in every instance a large drop when freed from the influence of general impression as measured by the total score. Halo effect is obviously the principal determinant of the ratings on single items. Following Symonds' reasoning (4), we might consider the difference between the average correlations as originally com-

TABLE 2

Net Correlations between Certain Items after Partialing Out the Total Score or General Impression

Items	r	Items	r
1– 9	.407	1–14	.344
1–13	.607	1–16	.161
2– 9	.501	2–12	.019
7– 8	.511	4–10	.309
7–10	.488	7–11	.238
9–13	.201	7–12	.524
11–12	.115	7–14	.442
Average	.449	Average	.097

puted for these two groups of item pairs and the average partial correlations shown in Table 2 to represent the amount of halo effect. Thus, the original average correlation of .715 for the first group of items is lowered to .449 when general impression is held constant; while the average r of .526 for the second group of items is lowered to .097. The common sense expectation of a low relationship between the pairs of items in the second group corresponds, therefore, to a net correlation of zero; while the *a priori* expectation of some overlapping between the pairs of items in the first group corresponds to a net correlation averaging .45.

This appears to mean that, although the correlation between the items is boosted by halo effect, the specific influence of the separate items is not wholly obscured. In other words, a speaker's average rating on the entire scale is the most dependable index of the rater's general impression. The speaker's average rating on each item reflects to a considerable degree this general rating. But, to some extent, the pattern of intercorrelations among items reflects a tendency toward analytical rating, *i.e.*, a tendency for the specific factors suggested by the item and its descriptive adjectives to influence the ratings on that item.

A profile of the speaker's average ratings on each item will thus show relatively little variation from one item to the next, or much less variation than if the audience were unaffected by this tendency to rate largely on the basis of general reaction. However, the items on which the speaker is most favorably rated probably represent quite validly his strong points, while those on which he has the lowest average ratings probably indicate the aspects of his performance which were actually least effective. Moreover, it is probable that the speaker really varied considerably *more* from the mean (or general impression level) of his own performance than the profile would indicate, and varied in the direction shown by the fluctuations of the profile.

An equally reasonable interpretation would emphasize the

practical significance of the dominant effect of general impression. Clearly, variations in the observable aspects of a speaker's behavior are to a great extent compressed toward a rating reflecting the audience's general reaction toward him. The variables which are presumed to enter into the total effect of a speech are certainly not perceived separately and distinctly, except to a limited extent which may give a clue as to the direction of their individual influence on the general impression. The important research problem, therefore, as well as its practical corollary, would be to discover what it is about a speaker that principally determines the audience's general impression.

To make possible a more exact comparison of data on the B–W Scale and the Purdue chart, two additional sets of calculations were made. Table 3 shows the reliability of the separate items on the B–W Scale, computed for the same group of ratings used in the previous calculations.

TABLE 3

Reliability of the Individual Items of the B–W Scale Based on the Average of Twenty Ratings

Item	r	Item	r
1	.62	9	.84
2	.91	10	.76
3	.90	11	.80
4	.87	12	.74
5	.93	13	.53
6	.73	14	.33
7	.85	15	.65
8	.88	16	.59

These figures for reliability of separate items were obtained by dividing the twenty raters for each speaker into two groups of ten each. To make them comparable to the inter-correlations on the fourteen pairs of items referred to above, the reliability of each item based on an average of twenty ratings was estimated by the Spearman-Brown formula. Besides noting the average reliability of ratings on individual items, we may

now arrange the data on these fourteen inter-correlations to show whether, as in the Purdue data, the inter-correlations exceed the square root of the product of the reliabilities of the two variables. This comparison is presented in Table 4.

TABLE 4

Inter-correlations between Pairs of Items Compared with Geometric Means of their Reliabilities

Items	r	Geom. mean of reliabilities	Items	r	Geom. mean of reliabilities
1–9	.63	.72	1–14	.59	.45
1–13	.74	.57	1–16	.37	.61
2–9	.77	.87	2–12	.50	.82
7–8	.75	.87	4–10	.47	.81
7–10	.78	.80	7–11	.75	.83
9–13	.70	.67	7–12	.44	.79
11–12	.59	.77	7–14	.48	.53
Average	.71		Average	.53	

The general trend of these figures seems fairly clear. The square root of the product of the reliabilities of the items exceeds the inter-correlation coefficient in 10 of the 14 cases. The average of these fourteen inter-correlations, computed by the use of Fisher's table of z scores, is .63. The average of the reliabilities of the twelve items involved in the inter-correlations, computed by the same method, is .77. The inter-correlations were determined by comparison of the average ratings of two audience samples, chosen at random. As noted above, Monroe, Remmers and Lyle found reliability coefficients of the order of .30, as against inter-correlations of the order of .60, based on the same subject's rating of two items. In their study, inter-correlation between items was still of the same order of magnitude as the reliability coefficients when one person's rating for a given item was correlated with a different person's rating for another item.

It is evident from the data in Table 4, most closely comparable perhaps to the inter-correlations obtained by Monroe,

Remmers and Lyle by the second of the two procedures just described, that ratings on the items of the B–W Scale do not correlate as highly as the reliability of the items permits. This reinforces the interpretation drawn from previously published data (1, Table II, p. 89), and from Tables 1 and 2, that the items of the scale possess varying degrees of independence from scores representing the rater's general impression and make correspondingly specific contributions to the total scores. There appears to be no question, however, that members of an audience react to a speaker predominantly in a general way, rather than analytically, and that further information useful in the prediction or control of audience reactions may be expected principally from a fuller understanding of the correlates of this general impression.

VARIABILITY OF SCORES

To secure some information about the variability of audience reaction, an analysis was made of approximately 1000 ratings obtained from two groups of raters: a group of 31 college sophomores at Pratt Institute, and a group of 25 college sophomores at New York University. Calculations were based of 40 speakers.

The range of scores for each speaker, the standard deviation of this distribution of scores, and the coefficient of variation were determined. For one college group, the average range of scores was 25 points on a scale of 64 points (1935 revision of the scale). The average V for the 25 speakers was 11.8. For the other college group, the average range was 18.3 and the V for this group was 9.7. In other words, a speaker may expect a range of reaction to his speech which covers a third of the entire gamut. On the basis of the S.D.'s of these distributions, one may expect an average speaker to receive a distribution of ratings with a sigma of 6 points on a scale of 64 points.

TIME OF RATING

The question arose whether rating tendencies might be affected by length of the time interval between hearing the

speech and rating the speaker on his performance. To throw light on this question, delayed ratings were obtained for comparison with ratings made immediately after hearing a speech. The two college groups referred to above made their ratings at the time of hearing the speeches. The third group, which had heard a series of one-hour lectures over a period of months, rated three of the speakers at the conclusion of the series. The intervals between the speeches and the time of rating were 3, 5 and 7 weeks.

Comparisons of these delayed ratings with those made by the other two groups showed (1) that the reliability of the delayed ratings was high and within the range found for the other groups, (2) that the average correlations between the two parts of the 1935 revision of the scale were lower than the average of the odd-even correlations, and (3) that the variability of the ratings was similar to that for the groups which had rated the speeches immediately.

EFFECT OF AGE OF RATERS

The effect of the rater's age upon his ratings of speakers was studied for the group of 36 speakers upon which data reported in Tables 1 to 4 was based. The age range in these audiences, 21 to 40 years, is sufficient to furnish a test of this factor for similar, fairly homogeneous audiences, although of course it does not represent the extremes which might be included if one were interested in comparing young with elderly persons. To learn whether the effect of age in such a group was fairly stable from speaker to speaker, correlations were computed for twelve different speakers, as shown in Table 5.

The correlations in Table 5 are based on 20 ratings for each speaker. The mean age of the raters was 26 and the S.D. of the age distribution was 4.8. For this group, the general trend of the correlations appears to indicate that there is no significant relation between B–W scores and the rater's age.

INFLUENCE OF SEX OF RATERS ON B–W SCORES

A comparison of ratings (on the 1935 scale) on the same speakers by the eighteen men and seven women sophomores of the New York University group showed an average score based on the men's ratings of 51.6 and on the women's ratings of 52.2. The correlation between the average score for each speaker based on men's ratings and the average score based on women's ratings was .82. The number of cases is of course too small for generalization, but in this comparison no important differences were found.

TABLE 5

Correlations between Age of Rater and Total Scores on B–W Scale

Speaker	r	Speaker	r
1	−.15	7	.14
2	.25	8	.34
3	−.03	9	.05
4	.06	10	.10
5	.63	11	.00
6	−.06	12	−.19

Although no audience situation was especially arranged to provide a large number of cases for a comparison of men's and women's ratings, information was available for several small groups. A college dean, speaking before a large audience of adults, received an average score of 60.9 from 18 men and 64 from 10 women. A prominent woman radio speaker was rated 56 and 59.5 respectively on one speech by 2 men and 8 women, and 59.7 and 55.6 on a second speech by 3 men and 14 women. (In rating radio speakers, items 2 and 8 on personal appearance and platform behavior are omitted.) A well-known male radio speaker was rated 58.3 and 58.9 by 3 male and 7 female listeners. As in the case of the earlier data, no greatly different average rating was obtained on any speaker, and no consistent tendency for a higher rating by either sex is evident.

It may be that comparisons based on more adequate samples will continue to show no general sex difference in rating tendency. This would not mean, of course, that all speakers or topics would be equally approved by men and women auditors, but would make possible a more exact description of situations in which such differences might occur.

INTELLIGENCE AND PERSONALITY OF SPEAKER

An attempt was made to study the relationship between the average B–W score on a speech and the intelligence and personality of the speaker as measured by standard tests. A group of art education students at Pratt Institute served as subjects for this experiment. Each was rated on a five-minute speech by 20 raters. Each student speaker was given *The American Council on Education Psychological Examination for College Freshmen*, 1938 edition and the *Bernreuter Personality Inventory*. In computing correlations, the composite language score (completion test, same-opposite test, and artificial language test) was used for the American Council Test; while the Bernreuter Inventory was scored for B_1N and F_2S.

For a group of 76 speakers, the correlation between intelligence (as measured by ability to deal with verbal symbols on a paper and pencil test) and average B–W score was .151 with a P.E. of .08. Correlation between average B–W score and Bernreuter B_1N score, based on 57 speakers, was .076 with a P.E. of .09. For a group of 42 speakers, the correlation between average B–W score and Bernreuter F_2S was .044 with a P.E. of .10. (Since Bernreuter B_1N and F_1C scores themselves are correlated to the extent of .95, it is evident that the average B–W scores would show about the same relationship to the latter as they do to the former measure.)

These findings indicate that, for this group, there was no reliable relationship between audience evaluation of a speech and the speaker's performance on standard tests of intelligence and personality. Similar findings have been reported by Gilkinson and Knower (5) and by Hayworth, Rondileau, Miller and others (6).

PLATFORM VERSUS RADIO RATINGS

It is apparent that items 2 (*personal appearance*) and 8 (*platform behavior*) are irrelevant to the rating of speeches broadcast by radio. The omission of these two items makes a change in the range and mid-point of the B–W scale. The 14-item scale used for rating radio speeches has a range of 14 to 70 points, with a mid-point at 42. While frequency curves drawn to the same scale can be compared if the mid-points for the two forms of the scale are superimposed, it may be that the absence of visual stimuli will affect the ratings more than by the simple arithmetical reduction of the average score in this proportion. For this reason the distribution of scores for a large number of radio speakers will be studied before any basis for comparison is suggested.

REFERENCES

1. Bryan, A. I., and Wilke, W. H. A technique for rating public speeches. *J. Consult Psychol.*, 1941, 5, 80–90.
2. Wilke, W. H. A speech profile. *Quart. J. Speech*, 1940, 26, 625–630.
3. Monroe, A. H. Remmers, H. H., and Venemann-Lyle, E. Measuring the effectiveness of public speech in a beginning course. *Studies in Higher Education*, 29, *Bull. Purdue Univ.*, 37, 1936. Pp. 29.
4. Symonds, P. M. Notes on rating. *J. Appl. Psychol.*, 1925, 7, 188–195.
5. Gilkinson, H., and Knower, F. H. *Psychological Studies of Individual Differences among Students of Speech.* Minneapolis, 1939, Pp. 196. Privately published.
6. Hayworth, D., Rondileau, A., Miller, R. A., and others. *A Research into the Teaching of Public Speaking.* Detroit, 1939, National Association Teachers of Speech. Pp. 233.

ERRORS IN PERCEIVING THE TEMPORAL ORDER OF AUDITORY AND VISUAL STIMULI

L. BALD, F. K. BERRIEN, J. B. PRICE, AND R. O. SPRAGUE*
Colgate University Psychological Laboratory

THE study reported in this paper concerns responses to visual and auditory stimuli occurring in rapid succession, and the factors affecting such responses. Practically this question is important because criminal cases frequently depend upon testimony concerning the time and order of fast moving events. Did the autos collide just before or after the traffic lights changed? Did the accused start running just before or just after the burglar alarm sounded? The guilt or innocence of defendants has depended upon accurate testimony on points such as these. Out of questions of this sort our study evolved.

From the psychological standpoint these questions reduce themselves to a very simple experimental problem: Within what time interval are individuals unable to distinguish the correct order of two successive stimuli such as a light and a sound? This problem has been the object of psychological interest for many years. The roots of the problem extend far back in the history of experimental psychology to studies of the personal equation among astronomers, and the complication experiment of the last century.

Many earlier studies have been criticized on the grounds of inadequate apparatus and uncontrolled experimental factors. Furthermore, the preceding experiments have not directed themselves to the practical aspects of questions such as we have attempted to answer. The present investigation concerned

* With assistance of R. T. Libby.

itself with three aspects of the problem: *first,* determining the presence or absence of a "predisposition" that would distort the apparent time order of successive impressions; *second,* the effect of the observer's position, relative to the stimuli, on the accuracy of his report; *third,* the influence of obscuring the sound source from view.

APPARATUS

The apparatus consisted of a constant speed motor driving a brush that made contact upon each revolution with two separate, adjustable poles. This brush was timed to make a complete revolution in one second. The poles were mounted above a dial divided into 100 equal units. Wired to one of these adjustable poles was a 10 watt neon bulb, capable of instantaneous brilliance. For the first portion of our study the second adjustable pole was connected to a 110 volt A.C. low-lag relay which gave a sharp click when contact was made. This relay was placed out of sight behind a screen that obscured from view the entire apparatus except for the neon bulb which was mounted with its round disks flat, rather than edgewise to the subjects. In the second portion of the experiment a magnetic speaker was substituted for the relay and was mounted in view just below the neon bulb.

The light and the sound stimuli in both experiments were judged to be approximately of equal intensity. This was primarily a precautionary measure since Smith's[1] experiment shows little effect of varying intensities on judgments of this kind.

PROCEDURE AND SUBJECTS

The apparatus was placed in a lecture room measuring 27 by 40 feet so as to allow clear vision for all subjects. The position of each subject was recorded. The group in both experiments consisted of an elementary class in psychology made up entirely of men. In the first experiment 32 subjects were present; in the second only 30. The room was treated

[1] W. F. Smith, "The Relative Quickness of Visual and Auditory Perceptions," *J. Exper. Psychol.*, 1933, 16, 239–257.

to reduce echoes. The subjects then were given record sheets with instructions to attempt to determine which came first, the light or sound, and a request was made that no trials be left

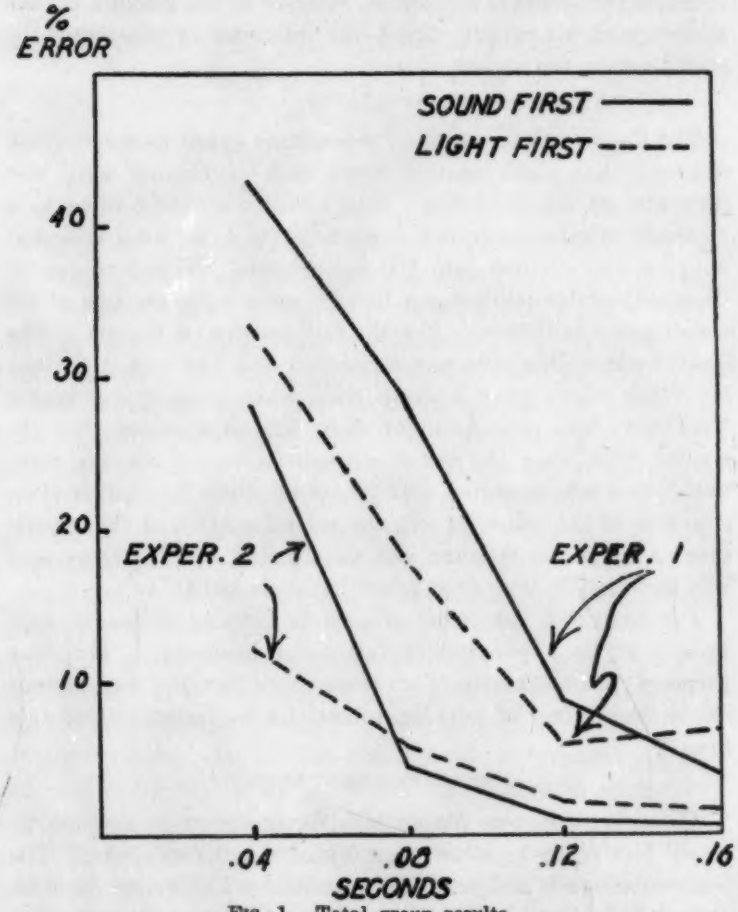


FIG. 1. Total group results.

blank. Care was taken not to emphasize in the instructions either light or sound. The experimenter made certain the subjects were ready before each trial by giving a warning signal. The two stimuli, light and sound, were varied in order

of presentation so no consistent pattern could be followed by the subjects. These two stimuli were, however, presented with different time intervals between them. The four intervals were .04, .08, .12, .16 seconds. Ten trials in hapazard order were given at each interval; 5 with light first, and 5 with sound first.

RESULTS

In Figure 1 the total results for both parts of the experiment are given. The ordinates of this figure represent the per cent of error while the abscissa represent the various time intervals between stimuli. An examination of this figure shows that in the experiment:

(1) Fewer errors occur in the second portion than in the first.
(2) At the shorter intervals, more errors are made in both parts of experiment when sound precedes light.
(3) As the time intervals between the two stimuli increase in duration, neither stimulus has a marked advantage over the other.

The data were further broken down into two groups of subjects; those that were in the less favored positions in the room to observe the stimuli and those who were in the center section directly in front of the apparatus. Twelve men constituted the "less favored" group and six were in the favored section. The results of these divisions of the class are presented in Figures 2 and 3. Inspection of these figures shows that men who have the best opportunity to observe, made as many or more errors when light rather than sound comes first. This contradicts the data for the entire group and also for the men in unfavorable positions. It is also well to note that at the smaller time intervals those individuals with the poorer opportunity to observe made more errors than those in the center of the class. However, the less favored students made fewer errors when sound preceded light than the central section of the class.

DISCUSSION OF RESULTS

In discussing the results of our work, several things come immediately to the fore. First, the major results of these experiments cannot be satisfactorily explained in terms of reaction time alone. It is an accepted fact that reaction to light is somewhat slower than to auditory stimuli. In other words we hear sounds more quickly than we see sights. If this in-

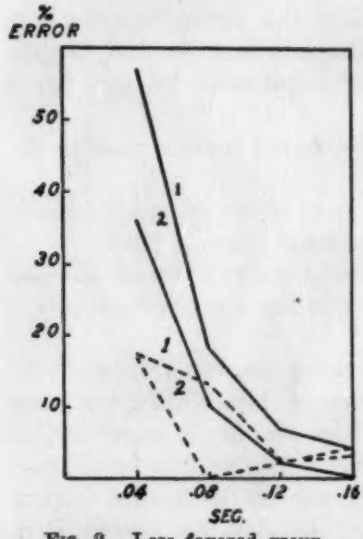


FIG. 2. Less favored group.

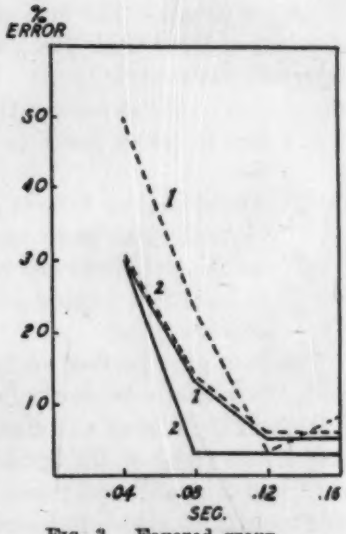


FIG. 3. Favored group.

formation is applied to the experimental situation we anticipate certain results. Since perception of sound is faster than perception of light, the psychological time interval is greater in the sound–flash order than in the flash–sound order. Therefore, we should, according to the reaction time explanation, expect the flash–sound order to give the greater error. The results recorded for the entire group are directly opposed to this expectation. The only scores we received which tallied with expectation were those of the subjects in the most favorable positions during the experiment. The major results in

Figure 1 are opposed to this expectation but are explained by an hypothesis originally set forth by Whipple[2] and later supported by Smith.[3] This hypothesis assumes that light has a greater attention-claiming quality, thus inducing the observer to report light as coming first in cases of doubt when actually sound preceded. It will be recalled for the first part of the experiment the apparatus was so arranged that only the light source was visible, the sound source being hidden behind the screen. In the second portion of the experiment we placed the sound source (a radio speaker) on the screen just below the light bulb. In this manner it was thought that the objective situation would "suggest" a more even distribution of attention between the two stimuli.

Figure 1 shows at the three longer intervals, the difference between the lower two curves is much less than the difference between the curves for the first experiment, thus lending good support to the hypothesis.

The marked lowering of both curves in the second portion of the study may, in fact, be attributed to a learning effect. Some incidental evidence is contrary to such an interpretation. During the preliminary trials of the apparatus, the investigators themselves made extensive observations as subjects. Learning effects were noted in the preliminary observations but were much smaller than the decrease noted in the group results. Consequently, at least a part of the differences observed between the two portions of the experiment are due to the altered stimulating conditions.

PRACTICAL IMPLICATIONS

The implications for evaluating testimony from this study are fairly obvious. It is evident that under conditions comparable to the classroom where attention is focused on the task, humans are surprisingly accurate in identifying the correct order of sights and sounds even when these events follow

[2] G. M. Whipple, "On Nearly Simultaneous Clicks and Flashes," *Amer. J. Psychol.*, 18, 98–99, 10, 280–286.

[3] *Op. cit.*

each other by split seconds. However, when the time interval between these events is less than a tenth of a second, errors in observation and description become numerous. Unless the witness is in a particularly favorable position to observe both the lights and the sounds, he is more likely to err in judging lights as preceding sounds, when really just the opposite occurs. In making any applications of this fact, one must bear in mind that the high degree of concentration and attention found in this experiment are seldom matched in the ordinary course of every day happenings, but the implications suggested may prove of some importance in evaluating testimony.

SUMMARY AND CONCLUSIONS

The present experiment has been designed to answer three questions: (1) Is there a predisposition to judge either light or sound as appearing first when the two stimuli follow each other in rapid succession? (2) What effect has the observer's position on his ability to make such judgments? (3) What is the effect of obscuring the sound source from view?

The judgments were made by a class in elementary psychology consisting of 30 to 32 college men in a large acoustically treated room. The data for the total group indicate: (a) that more errors of judgment occur when the order is sound–flash, than when the order is flash–sound at short intervals of time; (b) that obscuring the sound source from view increases the errors at the time intervals examined; (c) the best situated group made more errors when the order was flash–sound than when the order was sound–flash, while the poorly situated group conformed in general with the total group results.

The practical implications of this experiment have been developed.

NEWS AND NOTES

The Psychological Corporation offers for 1942–43 James McKeen Cattell Grants-in-Aid of Research in Applied Psychology including "Industrial, Personnel, Labor Relations, Social, Marketing, Advertising, Tests and Test Construction, Morale, Public Opinion, Propaganda, Clinical, and psychological aspects of related disciplines." The principal purpose of these grants is to add to research already under way by increasing the scope of the study or by permitting more intensive analysis of the data by students who have completed all or practically all course work for their doctoral degree in some field of applied psychology. The amount of each grant is $250.00. The closing date for the receipt of applications for 1942–43 on forms provided by the Grants-in-Aid Secretary is July 15, 1942. Further information may be obtained from the Secretary, Albert D. Freiberg, The Psychological Corporation, 522 Fifth Avenue, New York City.

It has been found necessary to change the date of the Conference on Human Development and Education, sponsored by the University of Chicago, the Commission on Teacher Education, and the Department of Supervisors and Directors of Instruction, from July 27 thru August 7 to August 10 thru August 21. The University is giving technical training to large groups of military personnel. Because these groups will occupy much of the dormitory space ordinarily used for workshops and conferences the change in time of meeting was made necessary.

A Conference on Human Development in War Time, sponsored by the Institute on Human Development of the Graduate School of Ohio State University, Columbus, Ohio is to be held from June 29–July 3, 1942. The purpose of the conference is to give those who have concern with the conservation of human resources an opportunity to extend their knowledge and understanding of human development and its relationship to the present emergency. Teachers, recreation workers, social workers, physical educators, nutritionists, sociologists, mental hygienists and parents will be interested in the Conference. In addition to a number of General Sessions, "special interests" groups will be organized to coordinate with the larger sessions. Information concerning accommodations, fees and registration may be obtained by writing to the Institute on Human Development, Graduate School, Ohio State University, Columbus, Ohio.

The purpose of the Institute on Professional Training for Clinical Psychologists, to be held at Pennsylvania State College from June 22 to 26, 1942, is to provide for a conference of all psychologists interested in the professional preparation of psychologists; to provide an opportunity for the Committee on Training in Professional (Clinical) Psychology, of the American Association of Applied Psychology, to hold meetings and have hearings of other psychologists interested in professional training; and to provide an advanced symposium of lectures and seminars for graduate students preparing for professional work in clinical psychology. Among the speakers on the program are M. R. Trabue, Edgar A. Doll, Carl R. Rogers, Percival M. Symonds, Marion A. Bills, Donald B. Lindsley, John G. Darley and others.

The Seventeenth Annual Meetings of the Midwestern Psychological Association were held at Hotel Statler, St. Louis, under the auspices of the Department of Psychology of Washington University, May 1 & 2, 1942. Dr. John P. Nafe, Chairman, and the other members of the department, were in charge of local arrangements. The formal sessions, arranged by Dr. W. N. Kellogg, included 54 papers on: comparative psychology, psychological tests, human learning, social psychology, abnormal psychology, conditioning, personality, and physiological psychology. Symposium topics included learning, conditioning theory, problems of graduate students, education and psychology, motivation, industrial and vocational psychology, propaganda and morale, and psychology and the war.

At the annual banquet Dr. E. A. Culler of the University of Rochester acted as toastmaster. Dean R. F. Jones of the Graduate School of Washington University, gave the address of welcome. The address of the retiring president, Dr. James P. Porter of Ohio University, was on the subject "Psychology and the Functional Integration of Human Behavior." Two hundred twenty-eight persons registered at the meetings, a wartime decline of about one-third as compared to other years at other cities similarly located.

New officers elected for the fiscal year beginning October 1, 1942, were: President: Robert H. Seashore, Northwestern University (1942–43); Secretary-Treasurer: Dael L. Wolfle, University of Chicago (1942–45). Members of Executive Council: Chester W. Darrow, Institute of Juvenile Research (1942–45) and Dean A. Worcester, University of Nebraska (1942–44) to fill unexpired term of former council member, Dael L. Wolfle.

The Association accepted the invitation of the University of Michigan to meet at Ann Arbor during the last weekend in April or the first weekend in May in 1943, unless governmental emergency plans interfere.

The June, 1942, issue of the Psychological Bulletin, edited by Arthur W. Melton, contains an interesting article entitled "Psychologists in Government Service" by Dr. Dael Wolfle, of the University of Chicago and a member of the Emergency Committee of the National Research Council. Believing that psychologists would be interested in knowing which ones of their colleagues are actively engaged in the services of the government, the former editor of the *Bulletin*, Dr. John A. McGeoch, Dr. Willard C. Olson and the author sent post-card questionnaires to each member of the A.P.A. Up to May 7, of the 370 questionnaires returned, 228 were from psychologists giving full-time service and 104 giving part-time service to the government. The following table shows the distribution of the 324 psychologists in the various branches of the service:

Social Security (including U. S. Employment Service)	61
Army (except for Air Force)	52
Army Air Force	48
Navy	38
Research for N.R.C. and N.D.R.C.	18
U. S. Public Health Service	18
Department of Agriculture	12
Civil Service Commission	12
U. S. Office of Education	7
National Youth Administration	7
Works Progress Administration	6
Office of Coordinator of Information	6
Federal Communications Commmission	4
Miscellaneous	35

A second table shows the distribution of these psychologists according to states, followed by a complete list of names and addresses so far compiled.

The author concludes: "Incomplete as these data are, they show unmistakably that psychologists from all parts of the country are serving their country in many capacities. When peaceful times make it possible to tell the whole story of the psychological work now going on, that story will be one in which many psychologists will have participated and of which all may be proud."

In addition to the above article appearing in the June, 1942 *Psychological Bulletin* are a number of others having to do with Psychology and the War. Dr. John G. Jenkins, now Lieutenant Commander A–V (S), U.S.N.R., has a contribution entitled "Utilization of Psychologists in the United States Navy," which falls under three headings: (1) Construction, administration, and interpretation of psychological tests. (2) Interviewing, varying from more or less typical pre-employment inter-

viewing through a variety of types to include clinical interviewing in search of signs of psychological abnormality. (3) Special services, including a wide variety of tasks for which the professional training of the psychologist makes him eligible.

Drs. L. A. Pennington and H. W. Case of the University of Illinois describe in their article the course in Military Psychology which they are supervising at the University of Illinois. "The Adjutant General's School and the Training of Psychological Personnel for the Army" is the title of an article by Morton A. Seidenfeld, Major, A.G.D. From this course Personnel Consultants are graduated and sent to duties in all types of Army installations, including work in the War Department as well as various types of field force units. The faculty for this course is a highly distinguished one including both military and civilian personnel. Among the civilian instructors are G. C. Garrett, Walter V. Bingham, Marion Richardson, C. S. Shartle, David Wechsler. The military faculty includes the following psychologists: Captain Donald Baier, Lt. T. W. Harrell, Lt. C. D. Leatherman and Major Seidenfeld.

"Tentative Suggestions on Undergraduate Psychological Training for Women in the Emergency," is the title of a report of the Subcommittee on Services of Women Psychologists in the Emergency prepared by Dr. Helen Peak and Dr. Edna Heidbreder.

The Ohio Association for Applied Psychology will hold its annual meeting on Saturday, June 20, 1942, at Ohio State University, Columbus, Ohio. It will open with a luncheon at the Dutch Tavern, High Street and 17th Avenue, for informal discussion. The officers for the coming year are President, O. A. Ohmann, Cleveland College; Secretary-Treasurer, M. W. Kuenzel, Children's Home, Cincinnati. In a business meeting of last year various committee chairman made reports. The secretary announced that sixteen members of the A.A.A.P. wished to join the Ohio Association.

Following the recent death of Dr. John A. McGeoch, the *Psychological Bulletin* will be published under the editorship of Dr. John E. Anderson, Institute of Child Welfare, University of Minnesota, Minneapolis, Minnesota.

Occupational Index, Inc., New York University, is announcing the publication of "Best Books of 1941 on Occupational Information, Guidance, and Personnel Administration," which is a list of 48 selected titles, eight of which are free. The others are arranged in suggested order of purchase for an average public school library.

BOOK REVIEWS

BENDER, I. E., IMUS, H. A., ROTHNEY, J. W. M., KEMPLE, C., AND ENGLAND, M. R. *Motivation and Visual Factors: Individual Studies of College Students.* Dartmouth College Publications, 1942. Hanover, New Hampshire.

In 1936 the Dartmouth Eye Institute began investigating the visual characteristics of the 1940 class. When statistical analysis revealed little relationship among visual conditions, scholastic performance and reading ability, further search was made for factors which obscured the relationships. If the "emotional blockages, contradictory but deep-seated traits of personality, or failure to appreciate the intellectual life often inhibit the educational advantages which might be expected to accrue to those who have no need of visual correction and to those whose visual defects have been adequately corrected by glasses," or "the high resolve, or the organization of personal values facilitates the student's performance, quite aside from the visual handicaps," then "the particular motivation in each case is significant. The problem thus concerns itself with the relevance of the visual factors for the motivation of the individual."

"Psychological investigation is being directed more and more to the problem of intra-individual consistency, as well as inter-individual uniformities; it will not be content with the discovery of laws pertaining to minds-in-general, but will seek also to understand the lawful tendencies of minds-in-particular." Individual case study is necessary to discover the motivating components which might be hidden in the data mass. From the entire class, 124 cases were selected for experimental study. The bases for selection were visual and educational factors. Of these, twenty representative cases were selected on the basis of visual factors for presentation in this report.

The data collected are subsumed under these heads:

Psychological aspects: interviews, projective techniques, attitudinal autobiography, personality questionnaires, behavior descriptions by members of the faculty and ratings by student associates.

Visual measurements: defects due to disease, corneal astigmatism, errors of refraction, ocular rotations, heterophoria, fusional amplitude, peripheral fusion, ocular dominance and perception of color. From the complete visual data a visual efficiency rating was given each subject. The visual defects ranked in influence upon visual efficiency as follows: near phoria, distant aniseikonia, refraction and near aniseikonia.

The Psychological Examination of the A.C.E., The Iowa Test of Silent Reading and The Graduate Record Examinations. Within tests or among grades the correlations are significant but are not significant between tests and grades. One of the highest r's is between freshmen grades and A.C.E., .59.

Medical and health records, extra-curricular activities, publicity records, administrative and personnel bureau records, and a follow-up questionnaire.

"A thorough examination and review of the protocol of case materials were made for each subject, and an abridged psychological case study was prepared. The psycho-portrait excerpts generously from the protocol and delineates the organization of the individual's traits, interests and values."

"The data are studied to determine his peculiar ways of reacting to the world within and the world without. The objective situation is represented by his interests, his behavior, his grades and the impressions he makes on others; the subjective point of view is derived from the thoughts, traits, purposes and values which he expresses directly or indirectly in the interviews, the projective techniques, personality questionnaires, etc. The organization of this psychological constitution is portrayed."

The twenty case studies include most of the book and make interesting reading. The psycho-portrait portion of a study exceeds in interest and space all other materials. No effort is made to hold certain traits as being responsible for low academic achievement.

The authors admit that the conclusions are restricted to the special population studied. Their approach to this problem has not been genetic, nor has there been any experimental control upon visual factors. Visual factors which interfere with visual comfort and efficiency should be corrected, but the data do not indicate the effect of visual defects upon the motivational pattern of the individual. But "the evidence indicates clearly that the motivational pattern of the individual does influence his adjustment to his visual condition. In every case the visual condition is so deeply imbedded within the total personality that its psychological significance is not clear . . . the lives of these college students are like the lives of most of humanity, replete with conflict and frustation."

An important question raised by this study respects "the values of piece meal studies of unit characteristics of groups." Two or even more measurements may indicate no correlation, reveal no causal agent. The individual must be considered as a whole. "Eye defects are part of the individual who has them."

Attention may here be called to some of the conclusions:

First, there is no statistical evidence of a significant relationship be-

tween visual efficiency and academic grades, or between eye defects and academic grades, or between eye defects and tested abilities.

Second, the subject's report of the severity and extent of his symptoms is more closely allied to his motivational pattern than to the degree of his visual defect.

Third. ". . . the particular motivational pattern of the individual is the principal factor that determines the nature of the individual's adjustment to his visual defects or their correction. Each student responds to his visual difficulty in a way consistent with his prevailing mode of adjustment. The visual condition cannot be considered apart from the motivational structure."

This study is very suggestive as to techniques, devices and problems concerning college students. The data are more evidence that we need to know the whole individual. The project which the Dartmouth Eye Institute has set for itself has far-reaching significance and the reports will be eagerly expected.

J. R. GENTRY,
Ohio University

ROETHLISBERGER, F. J. *Management and Morale.* Cambridge: Harvard University Press, 1941. Pp. xxii + 194.

WALTON, ALBERT. *The Fundamentals of Industrial Psychology.* New York: McGraw-Hill Book Co., 1941. Pp. xiii + 231.

The material furnishing the foundation for Management and Morale has been collected over a long period of time, but this is not apparent from simply reading it. It is evident from the title that it deals with organizations. Not alone business organizations as a whole but patterns of men within them; even going further and becoming concerned with organization within the individual. Indeed, the latter point turns out to be the principal one, when it is argued that the attitudes, standards of judgment, frames of reference of individual workers are things to be focussed upon and understood, if the manager is to maintain high morale.

The subject is sociologically oriented in the foreword written by Elton Mayo. The main body of the book consists of addresses given chiefly to business executive groups. The observations furnishing the impetus to the thinking set forth here were made in the Western Electric Company. But, the implications drawn from these studies, and apparently others, have here been woven together to make the most meaningful and complete explanation of why the constant struggle between managers and the men persists and recurs. And that is not all, for the judgments made from these observations suggest solutions to the trouble to which the author points convincingly.

It should be said that the difficulties shown and remedies suggested apply only for large industrial concerns and herein lies a tale.

The central idea is that men working together organize themselves into groups, aside from unions, for social as well as economic coöperation. Management in general has always recognized the economic organization, but ignored the social integration, except in small industrial concerns where it was possible for managers to make much the same kind of observations that furnish the foundations for this book. In large concerns managers have been so far away as to miss seeing this social organization among the workers. All kinds of complications result from this. Even the language of the two groups is different and communication on a meaningful basis cannot take place. The "economic" man cannot understand the "social" man. Communication between the social man of high status and his confrere of low status is stifled.

This book might have been entitled How to Deal with Coöperative Phenomena in Business, for only when the last chapter is reached is morale related to it.

It is difficult at first to reconcile the title of this book with its contents, but on further investigation it is seen that if methods can be found to generate coöperative phenomena in industry to replace compulsory measures, morale will be high. And, this would seem to be as true also on a still larger scale, for instance, in governments.

Books on personnel methods are numerous these days but it can be seen from the above that this one differs, to its advantage. Time and motion studies and the selection of men by tests avail one little if there is no way in which these people can gain enough satisfaction in the work situation to make them a relatively permanent part of it. Even though problems of selection are omitted, this is a book that is basic, along with Management and the Worker, and should be read by all who take the welfare of workers and their own businesses seriously to heart.

The second book, by Walton, is unlike the first in many respects. First, it is more comparable to the conventional book on psychology in industry. Second, it is less well documented. Third, it is better as a small reference book for people who have listened to lectures. It is designed to furnish points of departure for discussion of many phases of industrial psychology. Therefore, this book includes something on time and motion study, test construction and uses, learning, fatigue, motives, and attitudes. Some of these bits of information are derived from the Western Electric studies, but very few. It is easy to see how useful is this book, with its sixteen short chapters and 222 pages, for a group of tired business men in a one semester extension course. However, it is doubtful if reading this would make any change in the way they run their business. It is essentially a

survey and shows the ground that has been covered by psychologists in thinking about business. Probably business men would be interested by it, but there is less certainty of its being practical. The style is good and the book is easy reading, but its chief use could be only to furnish a starting point for a discussion which must be thoroughly comprehensive to be applicable.

Another way in which these books differ is that the whole of the first is devoted to the application of a single technique in industry. Although this makes for some repetition, it forces the reader to a single integrated conclusion, while the contents of the second are so heterogeneous that it suffers from being unable to "drive home" a single point.

Since both are directed toward the business man with little or no technical training in psychology, it seems better to the reviewer to offer him one good thing dressed up in several guises than to bring out the whole stock from which he can select the one he thinks he needs.

K. W. Oberlin,
University of Delaware

NEW BOOKS AND PAMPHLETS RECEIVED

Books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

Boletin do Services Social dos Menares. Vol. I, No. 2, December 1941. EDGAR SHALDERS, Director. Secretaria de Justice e Negocios de Interior. Sao Paulo, Brazil. 118 pp.

The Clarks, An American Phenomenon. WILLIAM D. MANGAM. Silver Bow Press, 289 Fourth Avenue, New York, 1941. $2.50. 242 pp.

The Child at Home and School. EDITH LEONARD, LILLIAN E. MILES and CATHERINE S. VANDER KAR. American Book Company 1942. 850 pp.

Employment Tests in Industry. A Selected, Annotated Bibliography. Industrial Relations Section. Princeton University, Princeton, New Jersey. 25 cents. 32 pp.

Handwriting Analysis. THEA STEIN LEWINSON and JOSEPH ZUBIN. King's Crown Press, New York, 1942. $2.00. 147 pp.

Intelligence in Mental Disorder. ANNE ROE and DAVID SHAKOW. Annals of the New York Academy of Sciences, Vol. XLII, January 1942. $1.00. 129 pp.

Motivation and Visual Factors. IRVING D. BENDER, HENRY A. IMUS, JOHN W. M. ROTHNEY, CAMILLA KEMPLE and MARY R. ENGLAND. Darthmouth College Publications, Hanover, N. H., 1942. 369 pp.

Psychology. FLOYD C. DOCKERAY. Prentice-Hall, New York, 1942. 504 pp.

Psychology of Adolescence. LUELLA COLE. Revised Edition. Farrar & Rinehart, New York, 1942. $3.25. 660 pp.

Psychology in Use. J. STANLEY GRAY and OTHERS. American Book Company, 1942. $3.25. 663 pp.

Radio and Television Bibliography. Sixth Edition. Columbia Broadcasting System, New York, May 1942. 96 pp.

A Research Memorandum on Methods of Studying Public Opinion in Relation to the War. FLOYD L. RUCH. Social Science Research Council, 230 Park Avenue, New York, 25 cents. 44 pp.

A Study in Judging the Opinions of Groups. R. M. W. TRAVERS. Archives of Psychology, Columbia University, New York, 1941. 73 pp.

Directory of American Psychological Periodicals

(Continued)

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $7.00 per volume of 500 pages. Two volumes every year. Acting Managing Editor, Roy M. Dorcus. Founded 1921.

Comparative Psychology Monographs—Baltimore: Williams & Wilkins Co. Subscription $6.00. 400 pages per volume. Roy M. Dorcus, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Provincetown, Mass.: The Journal Press. Subscription $7.00 per volume. One volume per year. Edited by Carl Murchison and an international cooperating board. Quarterly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

Psychological Abstracts—Ohio State University, Columbus, Ohio: American Psychological Association. Subscription $7.00 a year ($7.25 foreign). 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of Genetic Psychology—Provincetown, Mass.: The Journal Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Provincetown, Mass.: The Journal Press. Subscription $7.00. 500–600 pages annually. Edited by John Dewey, Carl Murchison, and an international cooperating board. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—New York: The Personnel Research Federation. Subscription $5.00. 400 pages. Edited by Charles S. Slocombe. Published monthly, except July and August.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Educational Psychology—Baltimore, Md.: Warwick & York. Subscription $6.00. Edited by Jack W. Dunlap, Harold E. Jones, and Percival M. Symonds. Published monthly except June to August. Founded 1910.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $3.00 per annum. Quarterly. Published simultaneously in Great Britain and the United States. Edited by Charles E. Spearman (London) and Karl Zener (Duke University). Founded 1932.

Journal of Psychology—Provincetown, Mass.: The Journal Press. Subscription $14.00 per annum; $7.00 per volume. 1000–1200 pages annually. Immediate publication for any accepted manuscript. Edited by Carl Murchison and an editorial board. Quarterly. The general field of psychology. Founded 1935.